



Title	Blood DNA virome associates with autoimmune diseases and COVID-19
Author(s)	Sasa, Noah; Kojima, Shohei; Koide, Rie et al.
Citation	Nature Genetics. 2025
Version Type	VoR
URL	<a href="https://hdl.handle.net/11094/99693">https://hdl.handle.net/11094/99693</a>
rights	This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

# Blood DNA virome associates with autoimmune diseases and COVID-19

Received: 24 October 2023

Accepted: 30 October 2024

Published online: 03 January 2025

 Check for updates

A list of authors and their affiliations appears at the end of the paper

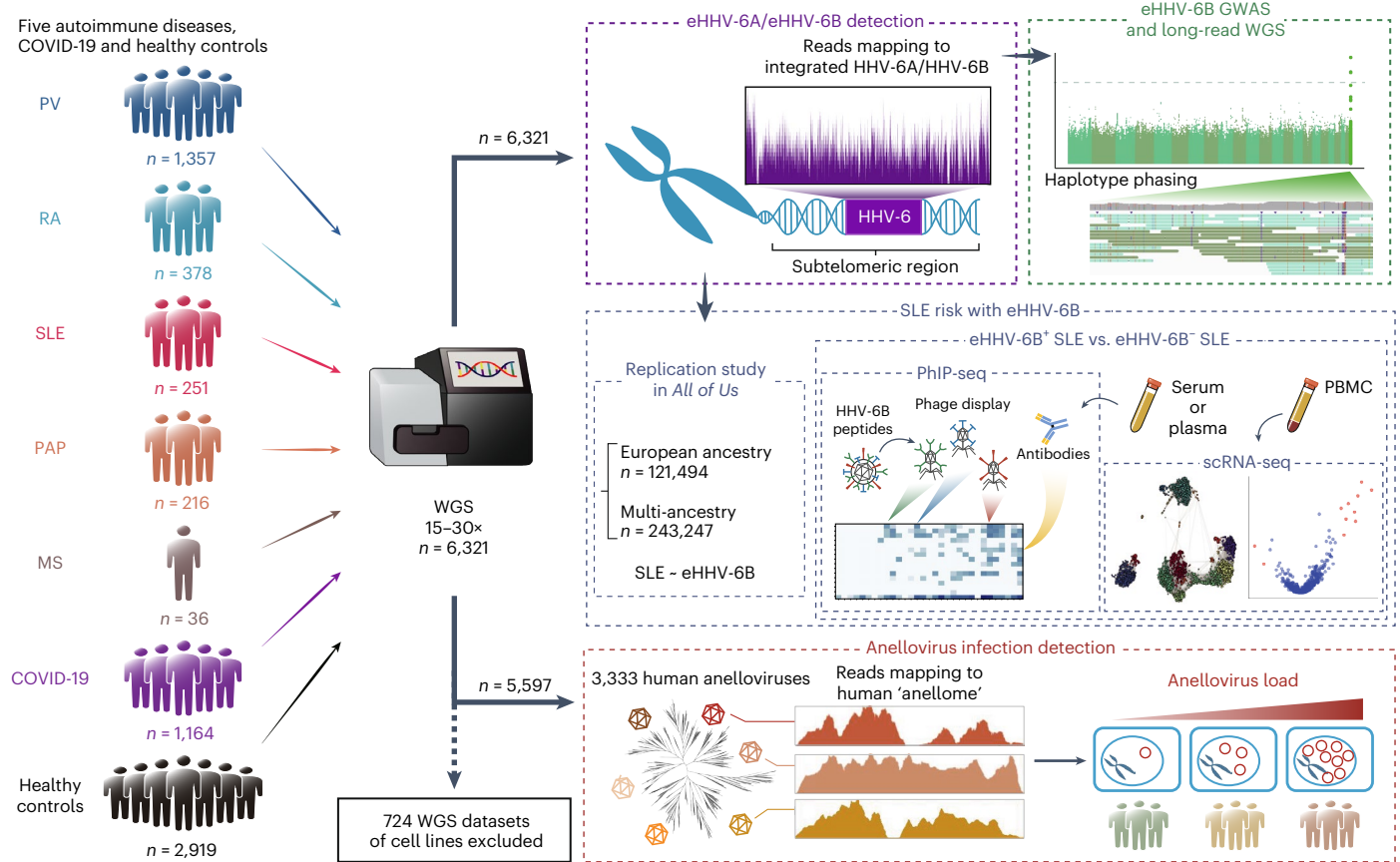
Aberrant immune responses to viral pathogens contribute to pathogenesis, but our understanding of pathological immune responses caused by viruses within the human virome, especially at a population scale, remains limited. We analyzed whole-genome sequencing datasets of 6,321 Japanese individuals, including patients with autoimmune diseases (psoriasis vulgaris, rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), pulmonary alveolar proteinosis (PAP) or multiple sclerosis) and coronavirus disease 2019 (COVID-19), or healthy controls. We systematically quantified two constituents of the blood DNA virome, endogenous HHV-6 (eHHV-6) and anellovirus. Participants with eHHV-6B had higher risks of SLE and PAP; the former was validated in All of Us. eHHV-6B-positivity and high SLE disease activity index scores had strong correlations. Genome-wide association study and long-read sequencing mapped the integration of the HHV-6B genome to a locus on chromosome 22q. Epitope mapping and single-cell RNA sequencing revealed distinctive immune induction by eHHV-6B in patients with SLE. In addition, high anellovirus load correlated strongly with SLE, RA and COVID-19 status. Our analyses unveil relationships between the human virome and autoimmune and infectious diseases.

Well-regulated induction of immune responses targeting nonself-viral epitopes is beneficial for virus clearance. On the other hand, excessive or aberrant responses to virus infection can resemble those of autoimmune diseases and occasionally pose risks of self-reactive immune responses as in Guillain–Barré syndrome<sup>1</sup>. Similarly, epitope indistinguishability between EBNA1, encoded by Epstein–Barr virus (EBV), and host GlialCAM likely contributes to the autoreactive lymphocyte expansion seen in multiple sclerosis (MS)<sup>2,3</sup>. In the mouse model, persistent infection by Theiler’s virus causes an autoimmune response targeting myelin via epitope spreading, where self-proteins are aberrantly presented to the immune system<sup>4</sup>. In systemic lupus erythematosus (SLE) plasma, antibody titers against an envelope protein encoded by endogenous retrovirus correlate with interferon (IFN)-stimulated genes (ISGs) expression<sup>5</sup>. In addition, the immunopathogenic response recognized as a ‘cytokine storm’ is attributable to both virus infection, as seen in severe coronavirus disease 2019 (COVID-19), and autoimmune diseases<sup>6</sup>. Our understanding of the relationships between virus

infection and autoimmune diseases is incomplete. For example, it is still unclear which autoimmune diseases are associated with virus infection and what factors license virus infection to initiate and exacerbate autoimmune disease.

DNA viruses often establish asymptomatic latent infections in healthy individuals. For example, EBV infection is highly prevalent, reaching more than 90% of adult individuals<sup>7</sup>. Anelloviruses are detectable in the blood of more than 8% of healthy people<sup>8</sup>. Beyond exogenous viruses, about 1% of humans carry endogenous HHV-6 (eHHV-6), a heritable HHV-6 genome integrated into the subtelomere region that has been proposed to be able to reactivate as a transmissible virus<sup>9,10</sup>. Because eHHV-6 is not included in reference human genomes, it is an overlooked genetic factor that may trigger antiviral responses. High-throughput sequencing has provided new methods to identify and study such viruses, characterizable as the human ‘virome’<sup>8</sup>. For example, whole-genome sequencing (WGS) of human specimens and cell lines has been used to study eHHV-6 prevalence and potential reactivation<sup>9,11,12</sup>.

✉ e-mail: [nicholas.parrish@riken.jp](mailto:nicholas.parrish@riken.jp); [yuki-okada@m.u-tokyo.ac.jp](mailto:yuki-okada@m.u-tokyo.ac.jp)



**Fig. 1 | Overview of the study design.** We collected 6,321 WGS datasets from patients with five autoimmune diseases (PV, RA, SLE, PAP and MS) and COVID-19 and healthy controls. The datasets were analyzed to detect eHHV-6 and anellovirus infection. In addition, we conducted a large-scale replication study of

the association between SLE and eHHV-6 using All of Us WGS data. We conducted GWAS of HHV-6B positivity. We performed long-read sequencing, PhIP-seq and scRNA-seq targeting eHHV-6B-positive patients with SLE (Methods).

Population-scale sequencing can suggest relationships between the presence of specific human viruses and diseases or immune responses. For instance, the presence of EBV nucleic acid in the spleen and blood is associated with an increased plasma cell number<sup>13</sup>. The presence of an ssDNA virus family discovered by virome profiling, called *Redondoviridae*, is associated with periodontitis<sup>14</sup>. Anellovirus load correlates with immunosuppression and acute respiratory disease<sup>15,16</sup>. The presence of eHHV-6 has been associated with pre-eclampsia, angina pectoris and increased levels of HHV-6-specific antibody titers<sup>17–19</sup>. These findings highlight the need for systematic virome analysis to identify new associations between viruses and human diseases. The partially revealed associations between viruses and autoimmune diseases, for example, between EBV and MS, suggest that similar relationships may be discoverable with further investigation.

In this study, we analyzed WGS from >6,000 Japanese individuals, including patients with five autoimmune diseases, COVID-19 or healthy controls. We systematically detected eHHV-6 and anellovirus nucleic acids in these datasets and conducted case–control association studies that compared eHHV-6 prevalence or anellovirus abundance in unrelated individuals with or without these diseases (Fig. 1). To confirm the inherited nature of the detected eHHV-6 genome, we performed genome-wide association study (GWAS) of eHHV-6B positivity and PacBio HiFi long-read sequencing of eHHV-6B-positive patients with SLE. In addition, epitope mapping by phage immunoprecipitation sequencing (PhIP-seq) from four eHHV-6B-positive and 12 eHHV-6B-negative patients with SLE and single-cell RNA sequencing (scRNA-seq) data from four eHHV-6B-positive and 11 eHHV-6B-negative patients with SLE were compared.

## Results

### Overview

We analyzed 6,321 Japanese samples. Details of the sample characteristics, sample collection and the WGS platform are described in Methods. The WGS datasets were composed of 1,357 participants diagnosed with psoriasis vulgaris (PV), 378 with rheumatoid arthritis (RA), 251 with SLE, 216 with pulmonary alveolar proteinosis (PAP), 36 with MS, 1,164 with COVID-19 and 2,919 healthy controls. The majority of the patients with COVID-19 ( $n = 1,068$ ; 91.8%) had severe disease (requiring oxygen support, artificial respiration and/or intensive care; Supplementary Table 1). Two bioinformatics pipelines were used to detect eHHV-6 and anellovirus infection (Fig. 1). Briefly, we remapped those WGS reads that did not map to the human reference genome to their respective viral reference genomes. For anellovirus detection, 3,333 human anellovirus sequences were used to accurately detect widely divergent anellovirus genomes. Logistic regression and Fisher's exact were used to calculate the adjusted odds ratio (OR) for the risk of each disease with eHHV-6 or three anellovirus infection states defined from normalized anellovirus viral load (viral copies per cell). In each disease with a high OR, we estimated the relationship between clinical metrics in each disease and eHHV-6 or three anellovirus infection states using linear regression or logistic regression and Fisher's test.

### eHHV-6B associates with autoimmune diseases and SLE activity

We identified eHHV-6A in 16 participants and eHHV-6B in 22 participants (0.25% and 0.35% prevalence, respectively; Table 1, Supplementary Table 2 and Fig. 2a). Of these participants, one healthy control

**Table 1 | Associations of eHHV-6A/eHHV-6B with autoimmune diseases and COVID-19 in Japanese individuals**

Disease	Number of total samples	eHHV-6A						eHHV-6B					
		No. (%)		Logistic regression			Fisher's test	No. (%)		Logistic regression			Fisher's test
		No.	(%)	OR	95% CI	Wald test <i>P</i>	Mid- <i>P</i> value	No.	(%)	OR	95% CI	Wald test <i>P</i>	Mid- <i>P</i> value
PV	1,357	2	0.15	0.46	0.07–1.93	0.34	0.22	6	0.44	2.2	0.62–7.47	0.22	0.20
RA	378	1	0.26	0.55	0.03–3.62	0.60	0.82	1	0.26	1.5	0.07–10.9	0.71	0.76
SLE	251	0	0.0	NA (quasi-complete separation)			0.40	4	1.6	6.9	1.61–27.2	0.0061 <sup>a</sup>	0.0060 <sup>a</sup>
PAP	216	0	0.0	NA (quasi-complete separation)			0.46	3	1.4	7.0	1.43–27.7	0.0075 <sup>a</sup>	0.022
MS	36	0	0.0	NA (quasi-complete separation)			0.87	1	2.8	10.9	0.55–71.6	0.034	0.085
COVID-19	1,164	2	0.17	0.61	0.09–2.50	0.54	0.31	1	0.086	0.56	0.03–3.77	0.61	0.46
Healthy controls	2,919	11	0.38	1.0	–	–	–	6	0.21	1.0	–	–	–
Overall	6,321	16	0.25	–	–	–	–	22	0.35	–	–	–	–

Wald test and Fisher's exact test were conducted as two-sided tests. <sup>a</sup>Wald test *P* value < 0.05/6 and adjusted OR > 1, or Fisher's test mid-*P* value < 0.05/6 and greater than healthy controls.

with eHHV-6A and four participants with eHHV-6B (one with PV, two with PAP and one healthy control) had evidence of the 'solo-DR' form of eHHV-6 (ref. 9).

Next, we tested for case–control associations between eHHV-6 and each disease using the logistic regression model and Fisher's test. No significant differences were observed in eHHV-6A prevalence in patients with any disease compared with healthy controls. In contrast, eHHV-6B was more prevalent in patients with SLE and PAP than in healthy controls (1.6% and 1.4% versus 0.21% prevalence; Fisher's test mid-*P* = 0.0060 and 0.022, respectively). After adjustment for age and sex, we confirmed eHHV-6B associations with SLE and PAP with an adjusted OR of 6.9 in SLE (95% confidence interval (CI) = 1.61–27.2; Wald test *P* = 0.0061) and 7.0 in PAP (95% CI = 1.43–27.7; *P* = 0.0075). These observations suggest that carrying eHHV-6B contributes to the risk of these diseases with a large effect size relative to previously known genetic and environmental factors of the diseases<sup>20–22</sup>.

To identify a possible association between eHHV-6B and clinical metrics of SLE or PAP, we used the linear or logistic regression model and Fisher's test (Supplementary Tables 1 and 3). eHHV-6B had no association with low serum complement levels, immunosuppressant use or prednisolone use in SLE (*P* = 0.46, 0.79 and 0.76, respectively). However, eHHV-6B had a significant positive correlation with high SLE disease activity index (SLEDAI) scores<sup>23</sup> (mean SLEDAI = 30.5 and 6.0 for eHHV-6B-positive and eHHV-6B-negative patients with SLE, respectively; *P* =  $1.3 \times 10^{-8}$ ). Three of four participants with SLE carrying eHHV-6B had very high SLEDAI scores ( $\geq 18$ ), suggesting that eHHV-6B may have a substantial clinical impact on disease activity (Fig. 2b). Of interest, three of four eHHV-6B-positive participants with SLE were positive for anti-RNP antibodies, which are typically seen in about 33% of patients with SLE<sup>24</sup>, while the fourth participant was positive for anti-Scl-70 antibodies, which is quite rare in SLE. Two eHHV-6B-positive patients with SLE developed complications from other autoimmune diseases. One patient who tested positive for anti-RNP antibodies had scleroderma and Sjögren's syndrome, and the other patient positive for anti-Scl-70 antibodies had scleroderma and antiphospholipid syndrome. Given these unusual clinical manifestations, inborn eHHV-6B positivity may predispose to atypical SLE pathogenesis. eHHV-6B was not associated with the titer of antigranulocyte-macrophage colony-stimulating factor autoantibodies in PAP (*P* = 0.90), which are considered central to the pathogenesis of the adult sporadic type of PAP<sup>25</sup>.

#### eHHV-6B also associates with high SLE risk in All of Us

We conducted an independent large-scale replication study of the association between SLE and eHHV-6 using WGS data from the National Institutes of Health's (NIH) All of Us program<sup>26</sup>. The participants in

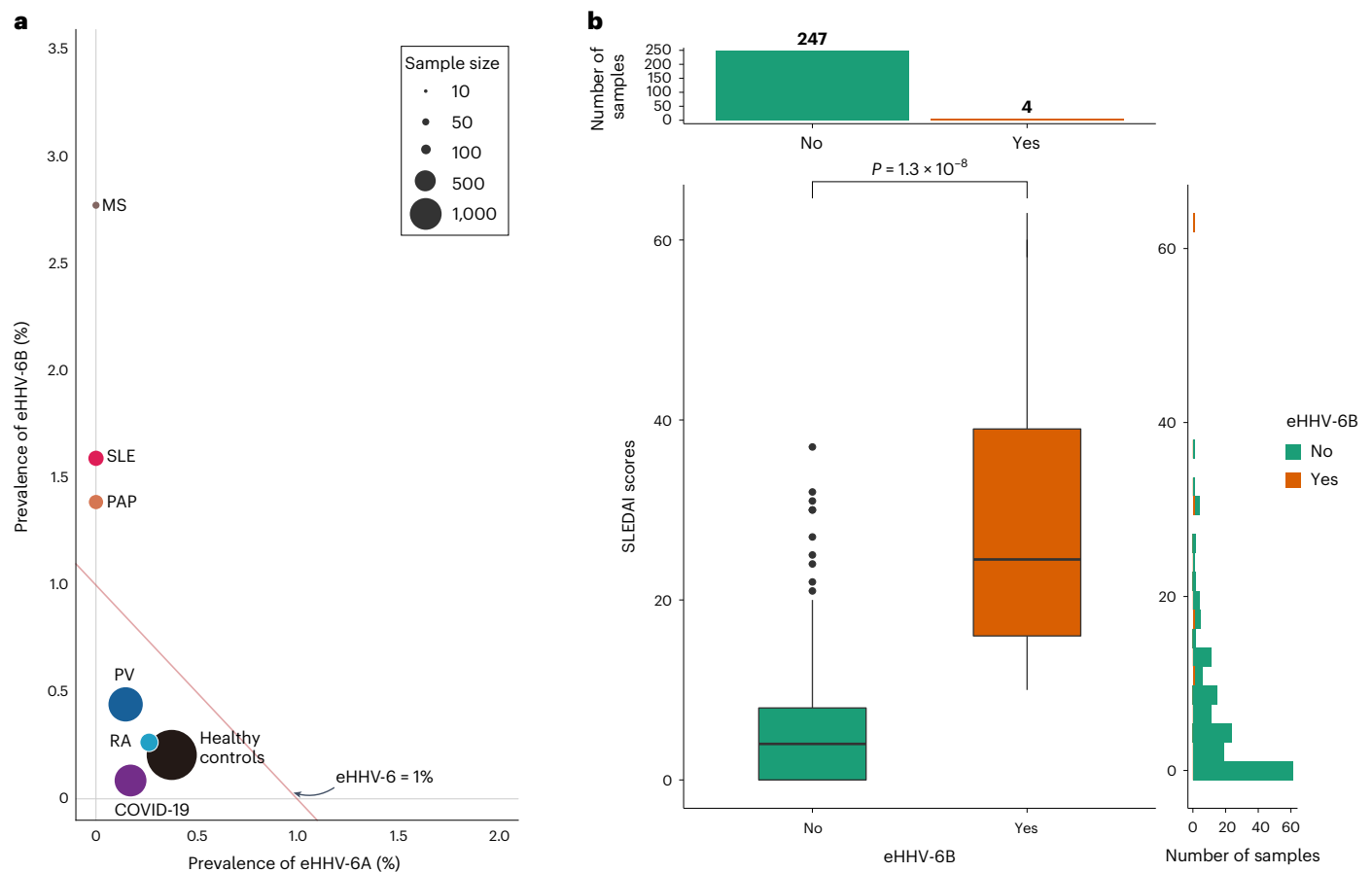
the All of Us cohort are genetically diverse, and the risk of SLE varies considerably by ancestry, so we performed association testing in the 121,494 individuals of European ancestry and the entire 243,247-person multi-ancestry dataset using a logistic regression model (Table 2). Concordant with the Japanese discovery cohort, no significant associations were observed in eHHV-6A prevalence in either population (*P* > 0.15). In contrast, eHHV-6B was associated with high SLE risk, with an adjusted OR of 2.7 in European ancestry (95% CI = 1.54–4.76; *P* = 0.0020) and 2.4 in multi-ancestry (95% CI = 1.47–3.82; *P* = 0.0012). These results indicate that eHHV-6B is associated with SLE in individuals of diverse ancestry.

#### Risk eHHV-6B of SLE locates on chromosome (chr) 22q

The sequencing depth of the HHV-6 genome detected in this study was nearly half that of the human genome in each participant, suggesting that the HHV-6 sequences are inherited along with single human haplotypes (Extended Data Fig. 1a,b). To understand the location of eHHV-6B in the human genome, we conducted a GWAS using common variants called the T2T-CHM13 reference. Comparison of 22 eHHV-6B-positive participants and 216 eHHV-6B-negative participants revealed a highly significant association with variants on the distal end of chr22q (Fig. 3a). This result was consistent with a previous GWAS using Fisher's test<sup>9</sup>, but using T2T-CHM13 as a reference provided a more densely mapped set of associated variants. Focusing on the distal end of chr22q, we performed logistic regression including rare variants (Fig. 3b,c and Supplementary Table 4). The 16 variants with *P* <  $1 \times 10^{-4}$  spanned a 500-kb region including different gene regions and different linkage disequilibrium (LD) block structures, but all were in high LD with eHHV-6B. By using 30× PacBio HiFi long-read sequencing, we confirmed that these tagged variants were on the same haplotype in each of the four eHHV-6B-positive patients with SLE (Fig. 3d and Extended Data Fig. 1c). These GWAS results reflect the LD relationship with the HHV-6B genome, suggesting that eHHV-6B in the Japanese population is integrated to the distal end of chr22q and has been inherited together as a linked unit. All of the tagged variants are intergenic or intronic, thus not supporting their pathogenicity, whereas all four eHHV-6B-positive patients with SLE carried the full-length HHV-6B genome implicated with viral reactivation<sup>27</sup> (Extended Data Fig. 1b). eHHV-6B itself rather than the tagged variants might therefore be involved in the pathogenesis of SLE.

#### eHHV-6B results in distinctive immunity in patients with SLE

To investigate whether eHHV-6 can result in the induction of unique antiviral responses, we performed epitope mapping of anti-HHV-6 antibodies in 16 patients with SLE (four eHHV-6B-positive and 12 eHHV-6B-negative; Supplementary Table 5) by PhIP-seq (Fig. 4a)<sup>28,29</sup>. As a result of epitope mapping to 108 HHV-6B proteins, two HHV-6B virion



**Fig. 2 | Prevalence of eHHV-6 and SLEDAI scores for eHHV-6B-positive patients with SLE. a**, Scatter plot of the prevalence of eHHV-6A versus eHHV-6B in participants with five autoimmune diseases, COVID-19 and healthy controls. The marker size indicates the sample size for each disease. The red line indicates that the prevalence of eHHV-6 = 1%. **b**, Box plot of SLEDAI scores for patients with SLE without and with eHHV-6B. SLEDAI scores in participants with eHHV-6B were significantly higher than those without eHHV-6B (mean SLEDAI = 30.5 and

6.0 for eHHV-6B-positive and eHHV-6B-negative patients with SLE, respectively; two-sided Wald test  $P = 1.3 \times 10^{-8}$ ). The histogram shows the distribution of the number of patients with SLE without and with eHHV-6B (top). The horizontal histogram shows the distribution of SLEDAI scores of patients with SLE without and with eHHV-6B, with different colors (right). Boxes denote the IQR; the median is shown as horizontal bars; whiskers extend to 1.5× IQR; outliers are shown as individual points. IQR, interquartile range.

**Table 2 | Replication studies of eHHV-6A/eHHV-6B associations with SLE in All of Us**

Disease	Frequency of the participants (%)	eHHV-6A				eHHV-6B			
		No.	OR	95% CI	Score test <i>P</i>	No.	OR	95% CI	Score test <i>P</i>
European ancestry ( <i>n</i> =121,494)									
SLE	<0.6	≤20 <sup>a</sup>	2.3	0.90–5.79	0.15	≤20 <sup>a</sup>	2.7	1.54–4.76	0.0020 <sup>b</sup>
Controls	>99.4	333	–	–	–	937	–	–	–
Multi-ancestry ( <i>n</i> =243,247)									
SLE	<0.7	≤20 <sup>a</sup>	1.4	0.54–3.51	0.68	≤20 <sup>a</sup>	2.4	1.47–3.82	0.0012 <sup>b</sup>
Controls	>99.3	502	–	–	–	1,578	–	–	–

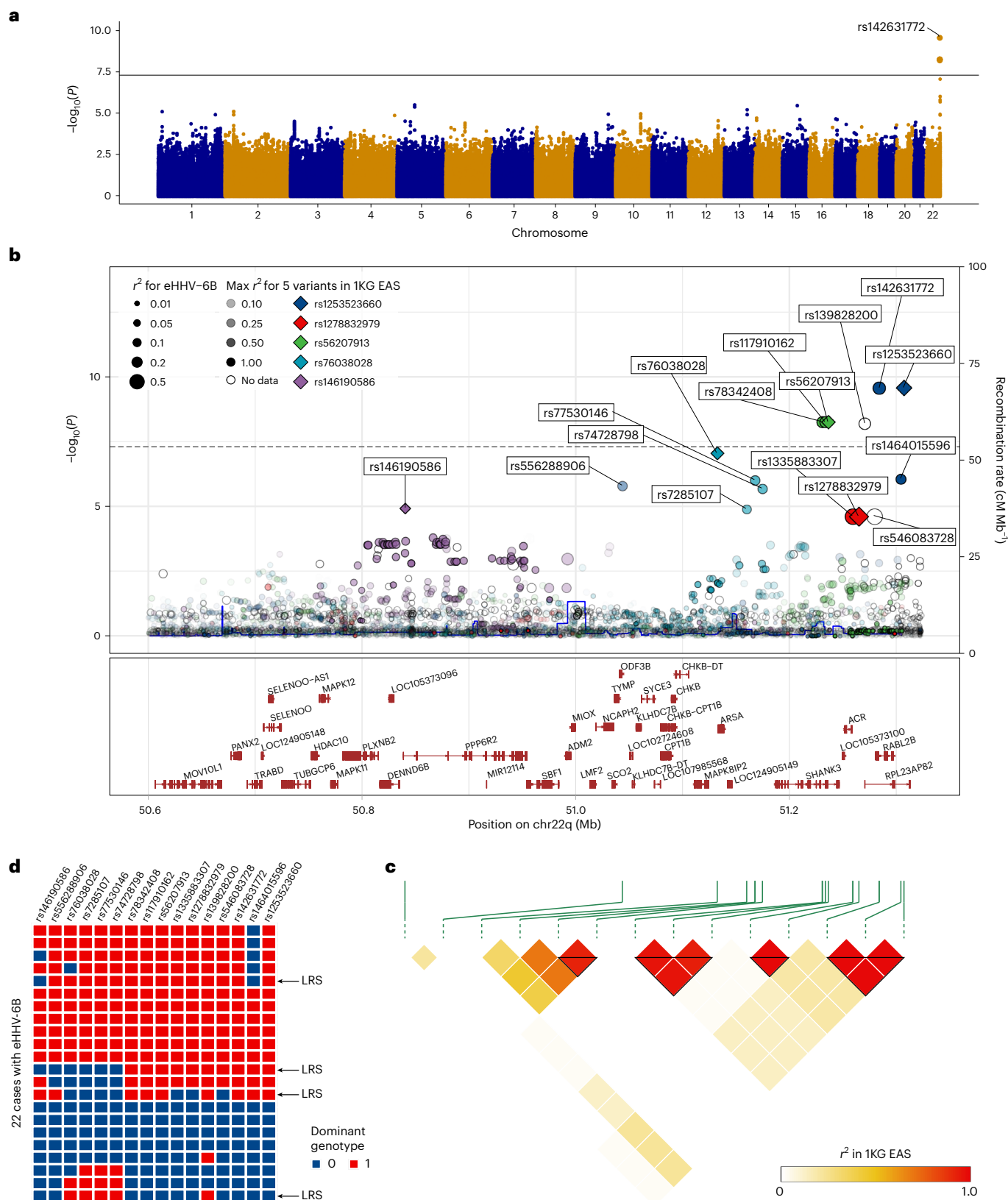
Score test was conducted as a two-sided test. <sup>a</sup>Counts suppressed per the All of Us Data and Statistics Dissemination Policy. <sup>b</sup>Score test *P* value <0.05/2 and an adjusted OR > 1.

proteins, antigenic virion protein and glycoprotein O, were frequently recognized as antigens across all participants, in agreement with the established notion that most adults have at one time been infected by exogenous HHV-6 and have acquired immunity against it<sup>30,31</sup>. To test whether there are epitopes specific to eHHV-6B carriers, we tested for an association between the presence of eHHV-6B and the antigenicity of HHV-6 peptides (Fig. 4b). Responses against one region (amino acids 476–490) in the immediate-early A (IE-A) transactivator were strongly detected in all eHHV-6B carriers, whereas responses against this region were not as strong in all eHHV-6 noncarriers ( $P = 3.5 \times 10^{-6}$ ; Fig. 4c).

This demonstrates that eHHV-6B results in distinctive immune responses against HHV-6B in patients with SLE. Because eHHV-6B originated as a virus but is now encoded in the human genome—it has the nature of both virus and self—eHHV-6B may be a heritable viral antigen to which immune responses represent autoimmunity, a phenomenon we term ‘endoimmunity’.

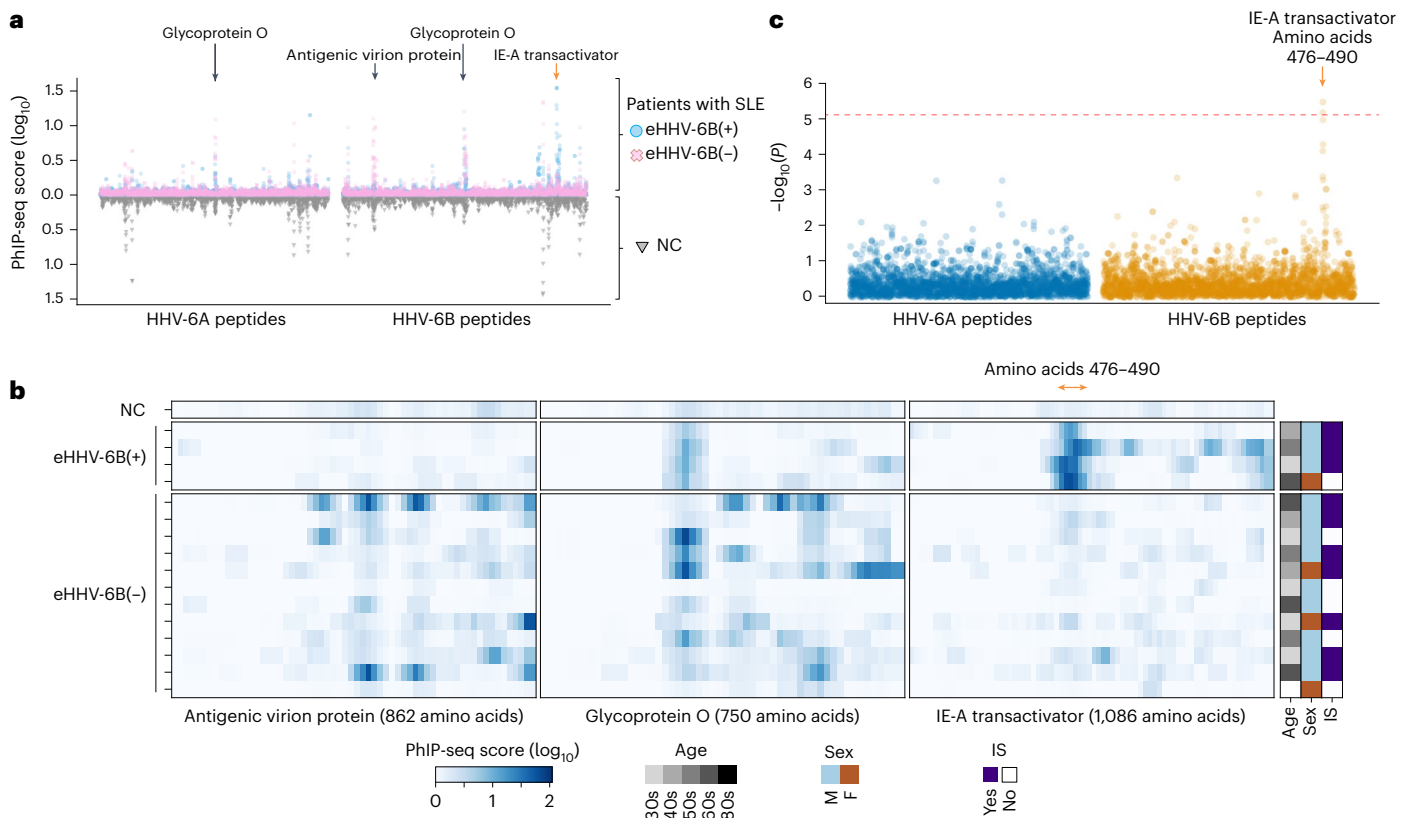
**eHHV-6B induces antiviral response in SLE monocytes**

Participants with both eHHV-6 and SLE showed evidence of distinctive humoral immunity to HHV-6 peptides. As SLE is characterized by



**Fig. 3 | eHHV-6B in Japanese integrated into chr22q. a**, Manhattan plot of common variant-based GWAS of eHHV-6B (22 cases versus 216 controls). The straight line indicates the genome-wide significance level ( $5 \times 10^{-8}$ ). **b**, Regional association plot for the distal end of chr22q. The dotted line indicates the genome-wide significance level. The 16 variants with  $P < 1 \times 10^{-4}$  were labeled by reference SNP ID (rs ID). Color shading represents the maximum value of LD ( $r^2$ ) with each of the lead variants in East Asian populations of 1KG. The circle size

indicates LD ( $r^2$ ) with the HHV-6B genome in the present WGS data. **c**, LD map of the lead variants in East Asian populations of 1KG. **d**, Heatmap representing the dominant genotype of the 16 variants in 22 eHHV-6B-positive participants. We confirmed that the variants were on the same haplotype in four patients with SLE with eHHV-6B by 30× HiFi long-read sequencing (Extended Data Fig. 1c). Uncorrected  $P$  values from the GWAS analysis (two-sided Wald test) are shown in **a** and **b**. LRS, long-read sequencing; EAS, East Asian.



**Fig. 4 | HHV-6 epitope mapping of antibodies in patients with SLE. a**, Detection of epitopes on HHV-6 proteins. Means of the PhIP-seq scores of the indicated participant groups are shown. Each dot shows a score for a peptide split by a 14 amino acids sliding window applied to HHV-6A and HHV-6B proteins. Specific proteins of interest are labeled. **b**, Detection of epitopes on the three HHV-6B proteins. Copy numbers of immunoprecipitated phage genomes encoding HHV-6B peptides are shown as a heatmap. Three HHV-6B proteins that harbor abundant epitopes are shown. The region of the IE-A transactivator associated

with eHHV-6B positivity is labeled with an orange arrow (amino acids 476–490). **c**, Statistical association between the presence of eHHV-6 and the copy numbers of immunoprecipitated phage genomes encoding HHV-6 peptides. IE-A transactivator amino acids 476–490 are labeled in orange. Uncorrected  $P$  values from the two-sided Wald test are shown. Red horizontal lines show the threshold of  $P(<0.05)$  after Bonferroni correction. IS, immunosuppressant use; NC, no antibody control; M, male; F, female.

immune cell dysfunction, we sought to further investigate the immunopathogenesis in eHHV-6B-positive participants with SLE. Peripheral blood mononuclear cells (PBMCs) were collected from the same individuals who were investigated by PhIP-seq (4 eHHV-6B-positive patients with SLE as cases and 11 eHHV-6B-negative patients with SLE as controls; Supplementary Table 5).

After processing using a unified single-cell analysis pipeline (Methods), we obtained 66,915 high-quality cells from PBMCs of 4 cases and 11 controls. There was no detectable expression of HHV-6B genes in any of the cells. Cells were annotated using Azimuth<sup>32</sup> (Fig. 5a). To identify more differential expression (DE) patterns than the pseudobulk methods or the mixed models, we applied Milo<sup>33</sup> and identified 2,818 neighborhoods, none of which showed evidence of differential abundance (DA; spatial false discovery rate (FDR) < 0.05), and 20 neighborhood groups consisting of neighborhoods that shared at least five cells (Fig. 5b and Extended Data Fig. 2a). After excluding several cell groups to avoid sampling bias, DE and Gene Ontology (GO) enrichment analyses were performed (Fig. 5c and Extended Data Fig. 2b–e). GO analyses showed that antiviral responses were enriched in cells in groups 6, 7 and 18, which were annotated as monocytes and dendritic cells (DCs). To evaluate the IFN response to viruses, we calculated antiviral ISG scores of cells in each neighborhood and compared them between cases and controls. In groups 6, 7 and 18, the antiviral ISG scores of the cases were significantly higher than those of the controls (FDR < 0.05; Fig. 5d). Within-group comparisons showed that the mean scores in the cases were significantly higher than those in the controls in groups 6, 7 and 18,

and the differences were larger than those in the other groups (Fig. 5e). Differences in the antiviral ISGs scores described above were observed even between eHHV-6B-positive and eHHV-6B-negative patients with immunosuppressant use, suggesting that these increased antiviral responses in monocytes and DCs were not attributable to immunosuppressants (Extended Data Fig. 3a,b). These results are consistent with the distinctive immunity to HHV-6B between eHHV-6B carriers and noncarriers revealed by the epitope mapping.

#### Anellovirus associates with autoimmune diseases and COVID-19

Anellovirus was analyzed in 5,597 Japanese WGS samples, excluding participants whose WGS was conducted using DNA extracted from cell lines. The anellovirus load was estimated and stratified (Methods). Of the 5,597 participants, 689 anellovirus-positive participants (12.3%; that is, those with anellovirus load above the detection limit), 57 participants (1.0%) with anellovirus load >1.0 considered anellovirus-infected and 16 participants (0.29%) with anellovirus load >8.0 considered anellovirus-viremic were found (Table 3, Supplementary Table 6 and Extended Data Fig. 4).

Correlations between the three anellovirus detection states and each disease were assessed using the logistic regression model and Fisher's test. The proportions of anellovirus infection were higher in patients with SLE, RA and COVID-19 than in healthy controls (4.8%, 3.7% and 1.7% versus 0.16%; mid- $P = 3.6 \times 10^{-10}$ ,  $7.4 \times 10^{-10}$  and  $2.9 \times 10^{-7}$ , respectively), and the proportions of anellovirus viremia were higher in patients with SLE, RA and COVID-19 than in healthy

controls (2.4%, 0.79% and 0.61% versus 0%; mid- $P = 5.6 \times 10^{-7}$ , 0.0023 and  $3.3 \times 10^{-4}$ , respectively). Anellovirus infection associated positively with SLE, RA and COVID-19, with an adjusted OR of 58.5 in SLE (95% CI = 15.6–293;  $P = 2.6 \times 10^{-8}$ ), 15.4 in RA (95% CI = 4.89–68.2;  $P = 2.5 \times 10^{-5}$ ) and 6.2 in COVID-19 (95% CI = 2.04–26.7;  $P = 0.0042$ ). Again, the quantitative magnitudes of the observed OR were larger than those previously reported for genetic and environmental factors of the diseases.

Next, linear or logistic regression models and Fisher's test were used to identify a possible association between anellovirus detection states and clinical metrics of SLE, RA and COVID-19 (Supplementary Tables 1 and 3). In SLE, anellovirus was positively associated with low serum complement levels, immunosuppressant use and prednisolone use, with an adjusted OR of 4.4 in participants with low serum complement levels (95% CI = 1.46–19.5,  $P = 0.020$ ), 3.0 in participants treated with immunosuppressants (95% CI = 1.51–6.03,  $P = 0.0020$ ) and 2.2 in participants treated with prednisolone (95% CI = 1.05–5.21,  $P = 0.047$ ; Extended Data Fig. 5a–d). Anellovirus detection states were not associated with SLEDAI scores. Although anellovirus detection states were insignificantly associated with rheumatoid factor (RF) or anticitrullinated protein antibody (ACPA), all participants with RA who were viremic for anellovirus ( $n = 3$ ) were RF- and ACPA-positive. Patients with severe COVID-19 ( $n = 1,068$ ) were not statistically at higher risk of anellovirus infection, but all anellovirus-infected participants with COVID-19 ( $n = 20$ ) had severe disease. The proportion of participants with anellovirus infection was statistically higher in participants who died of COVID-19 than in those who survived (mid- $P = 0.033$ ).

## Discussion

Here 6,321 Japanese WGS datasets were used to test for associations between six immune-mediated diseases and two constituents of the human blood virome, eHHV-6 and anellovirus. We present robust evidence of positive associations between these viruses and several autoimmune diseases and COVID-19. The observed risks of the diseases were relatively larger than those of the previously known genetic and environmental factors of the diseases, although their prevalence remains rare, suggesting that these viruses substantially impact disease biology and are relevant to clinical medicine.

We used WGS data, a robust method for detecting eHHV-6 (ref. 11), to estimate the higher prevalence of eHHV-6B in patients with SLE and PAP. eHHV-6 prevalence among healthy controls was 0.59%, similar to those in previous studies of healthy Japanese participants<sup>34,35</sup>. Our previous survey of eHHV-6 in a disease-oriented Japanese biobank found a higher prevalence of HHV-6B than HHV-6A, the opposite of that we observed here using healthy controls<sup>9</sup>. This is consistent with the present finding that eHHV-6B, but not eHHV-6A, is associated with an increased risk of several diseases and should motivate additional biobank-scale studies. It is also consistent with the independent replication data from the All of Us cohort.

Multiple facets of our analysis support a meaningful association between SLE and eHHV-6B. Patients with SLE with eHHV-6B had higher

SLEDAI scores, suggesting a role for HHV-6B in SLE pathogenesis and clinical activity, and half of them had other autoimmune complications. We identified the presence of distinctive immune responses against HHV-6B in eHHV-6B carriers with SLE. These findings are consistent with previous studies, reporting that people with eHHV-6 have higher antibody levels against HHV-6-encoded protein IE1 (ref. 19) and show different adaptive immune responses to other herpesviruses and immune gene expression patterns<sup>36</sup>. It is possible that the antibodies against the IE-A transactivator can act as autoantibodies; however, we cannot exclude the possibility that these are a consequence of broader immune activation against exogenous HHV-6 aided or triggered by eHHV-6B. Herpesviruses often encode immunomodulatory viral chemokine receptors<sup>37</sup>, and HHV-6 encodes the chemokine receptors U12 and U51 (ref. 38). The expression of these viral proteins may also contribute to pathological immune activation by eHHV-6B. A previous study found an association between SLE and circulating HHV-6 infection<sup>39</sup>. Single-cell analysis of PBMCs showed increased antiviral responses in monocytes and DCs. scRNA-seq data showed no eHHV-6 expression in PBMCs; however, this is far from mutually exclusive with the activation of humoral immunity against eHHV-6B we observed. eHHV-6 genes are weakly expressed in multiple human tissues other than blood cells, such as the brain and esophagus in the genotype-tissue expression, which recruited individuals without respect to specific diseases<sup>19</sup>. It may be possible that eHHV-6B carriers of SLE acquired antibodies to the HHV-6B protein after the onset of SLE. Although we cannot identify the timing of B cell activation against IE-A transactivator from the present dataset, it may be possible that eHHV-6B is one factor contributing to severity, rather than the initiating step of SLE. Further investigation into the 'endoimmunity' hypothesis as well as T cell activation characteristic of eHHV-6B carriers will shed additional light on the immunopathogenicity of eHHV-6B in SLE.

The variants in high LD with eHHV-6B found by the comprehensive analysis using T2T-CHM13 were unlikely to be pathogenic for SLE. Based on the present associations alone, especially in a single population, we cannot rule out the possibility that a variant that is in high LD with eHHV-6B is the mutation causative of increased SLE risk. However, eHHV-6B found in the Japanese and European populations is inserted in different chromosomes<sup>9</sup>. This implies that the variants in high LD with eHHV-6B in the Japanese population are different from those in the European population. Considering that eHHV-6B is associated with SLE in both populations, it is unlikely that a variant in high LD with eHHV-6B has a causal role in SLE. Furthermore, eHHV-6B is a polymorphism that has been overlooked in previous GWASs because it is not in the reference genome, complementarily enhancing our knowledge of human disease risk variants.

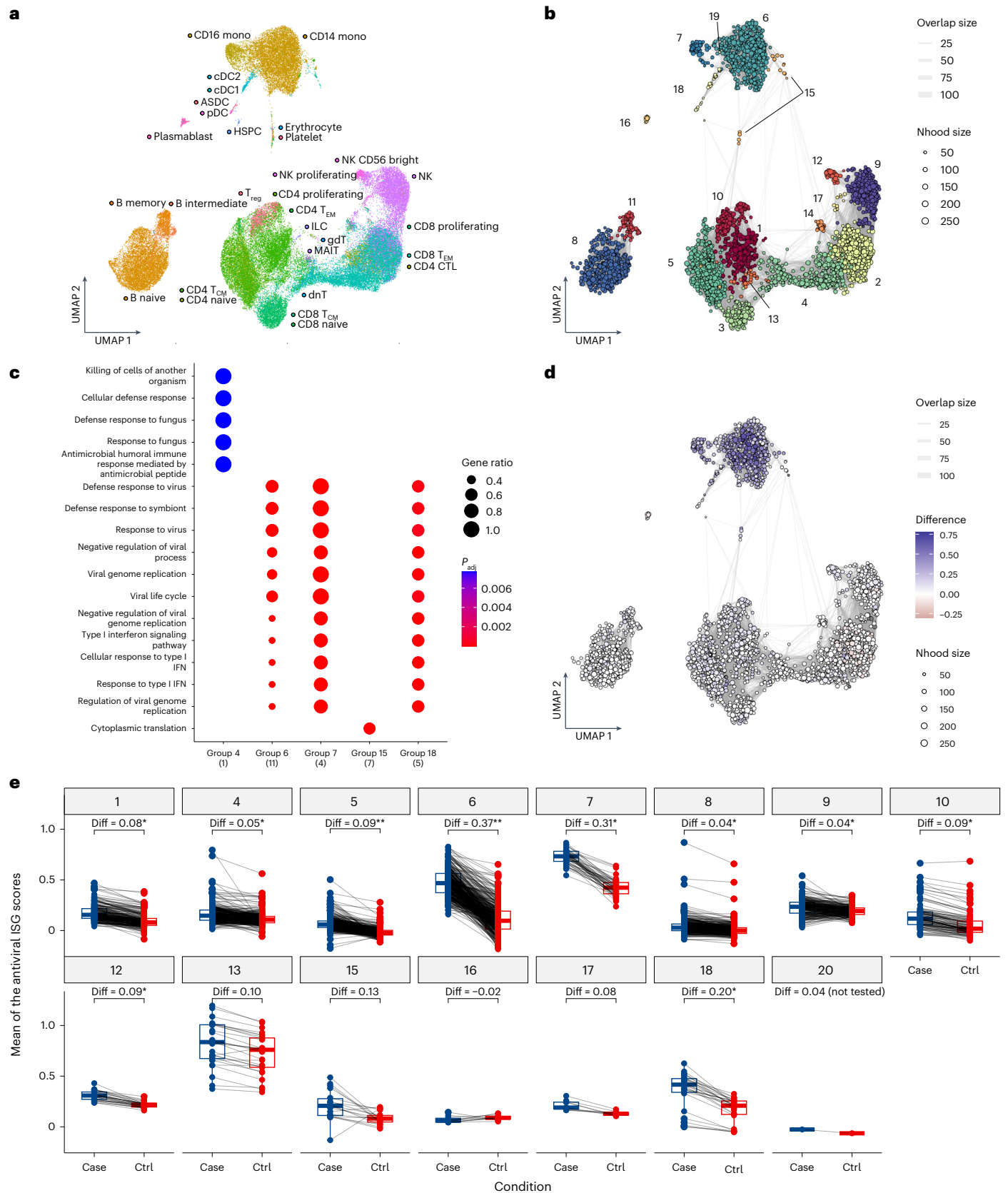
Anelloviruses are characterized by enormous genetic diversity<sup>40–42</sup>, which must be considered when detecting and estimating their viral load from WGS data. Here we did so by comprehensively using more than 3,000 anellovirus reference sequences. However, as with any mapping-based approach, we are limited by the degree to which

**Fig. 5** | scRNA-seq of PBMCs from eHHV-6B-positive and eHHV-6B-negative patients with SLE. **a**, UMAP embedding of scRNA-seq data for all 66,915 PBMCs from the participants with SLE ( $n = 4$  for eHHV-6B-positive and 11 for eHHV-6B-negative). Thirty cell types were defined by RNA expression of marker genes using Azimuth. **b**, Graphical representation of neighborhoods identified by Milo. Nodes are neighborhoods, colored by neighborhood groups, and their sizes correspond to the number of cells in each neighborhood. Graph edges depict the number of cells shared between adjacent neighborhoods. **c**, Enriched biological processes by GO analysis of DE genes of each neighborhood group. The total number of DE genes in each group is indicated in brackets. Dot color indicates the statistical significance of the enrichment (adjusted  $P$  values via the Benjamini–Hochberg method), and dot size represents the gene ratio annotated to each term. **d**, Graphical representation of neighborhoods identified by Milo as in **b**. Node color indicates the difference between the mean of the antiviral ISG scores

of cases and the mean of the scores of controls within the neighborhood (spatial FDR < 0.05), comparing eHHV-6B-positive patients with SLE as cases ( $n = 4$ ) and eHHV-6B-negative patients with SLE as controls ( $n = 11$ ). Neighborhoods with nonsignificant differences (spatial FDR  $\geq$  0.05) are colored white. Neighborhoods that contain only cells of either cases or controls are colored black. **e**, Comparison of differences within each neighborhood group shown in **d** using a two-sided paired  $t$  test. Diff is the average difference between the case and control means. Boxes denote the IQR; the median is shown as horizontal bars; whiskers extend to 1.5 times the IQR. \*FDR <  $1.0 \times 10^{-10}$  and \*\*FDR <  $1.0 \times 10^{-50}$ . ASDC, Axl+ Siglec-6+ dendritic cell; pDC, plasmacytoid dendritic cell; HPSC, hematopoietic stem and progenitor cell; T<sub>reg</sub>, regulatory T cell; NK, natural killer; T<sub>EM</sub>, effector memory T cell; ILC, innate lymphoid cell; MAIT, mucosal-associated invariant T; GDT,  $\gamma\delta$  T cells; CTL, cytotoxic T cell;  $P_{adj}$ , adjusted  $P$  value; Nhood, neighborhood.

these reference genomes represent the viral diversity in this population. Unlike germline integration of HHV-6, anellovirus status in the patient is not expected to match that of a cell line derived from the patient. Thus, 5,597 Japanese blood WGS samples were analyzed, and

the proportions of three continuous anellovirus detection states were assessed. Elevated anellovirus load was frequently found in the blood of participants with SLE, RA and COVID-19. Interestingly, patients with COVID-19 were anellovirus-positive in a similar proportion as healthy



**Table 3 | Associations of anellovirus infection and viremia with autoimmune diseases and COVID-19 in Japanese individuals**

Disease	Number of total samples	Anellovirus infection						Anellovirus viremia					
		No.	(%)	Logistic regression			Fisher's test	No.	(%)	Logistic regression			Fisher's test
				OR	95% CI	Wald test <i>P</i>	Mid- <i>P</i> value			OR	95% CI	Wald test <i>P</i>	Mid- <i>P</i> value
PV	1,062	7	0.66	4.0	1.01–19.8	0.061	0.025	0	0.0	–	–	–	1.0
RA	378	14	3.7	15.4	4.89–68.2	2.5×10 <sup>-5a</sup>	7.4×10 <sup>-10a</sup>	3	0.79	NA (quasi-complete separation)			0.0023 <sup>a</sup>
SLE	251	12	4.8	58.5	15.6–293	2.6×10 <sup>-8a</sup>	3.6×10 <sup>-10a</sup>	6	2.4	NA (quasi-complete separation)			5.6×10 <sup>-7a</sup>
PAP	216	0	0.0	NA (quasi-complete separation)			0.72	0	0.0	–	–	–	1.0
MS	36	0	0.0	NA (quasi-complete separation)			0.94	0	0.0	–	–	–	1.0
COVID-19	1,164	20	1.7	6.2	2.04–26.7	0.0042 <sup>a</sup>	2.9×10 <sup>-7a</sup>	7	0.60	NA (quasi-complete separation)			3.3×10 <sup>-4a</sup>
Healthy controls	2,490	4	0.16	1.0	–	–	–	0	0.0	1.0	–	–	–
Overall	5,597	57	1.0	–	–	–	–	16	0.29	–	–	–	–

Anellovirus infection, participants with anellovirus load >1.0; anellovirus viremia, participants with anellovirus load >8.0. <sup>a</sup>Wald test *P* value <0.05/6 and an adjusted OR >1 or Fisher's test mid-*P* value <0.05/6 and greater than healthy controls.

controls but included significantly higher anellovirus-infected and anellovirus-viremic than healthy controls. A possible scenario is that the atypical immune states characteristic of these diseases (or treatments) allow anelloviruses to reach higher levels in the blood<sup>43,44</sup>. This would also be supported by the observed association between anellovirus positivity and immunosuppressant and/or prednisolone use in SLE. An association between anellovirus positivity and low complement levels in SLE suggests that, similar to other viral infections, complement deficiency predisposes patients with SLE to anellovirus infection<sup>45,46</sup>. It is also possible that higher anellovirus levels contribute to the immune responses and inflammation leading to these diagnoses. Previous studies have shown that anellovirus DNA can induce proinflammatory cytokines<sup>47,48</sup>, suggesting that the higher load of anellovirus is a driver of severe inflammation during acute respiratory virus infection and, potentially, autoimmune response. While high-quality evidence of a causal relationship between anellovirus and human disease is limited, clonally expanded CD4<sup>+</sup> T cells recognizing an anellovirus, torque teno virus, have been isolated from the cerebrospinal fluid of a patient with MS<sup>49</sup>. Notably, our conclusion is similar to that of ref. 50 that the prevalence of detectable (at any level) torque teno virus DNA is not higher in patients with SLE or RA than in healthy controls. We did observe, however, that participants with SLE and RA are more likely to have elevated levels of anellovirus DNA, supporting the hypothesis that anellovirus replication may be involved in autoimmune diseases. Longitudinal rather than cross-sectional studies are needed to adequately assess this. Notably, in our results, the increased number of anellovirus-infected participants among those with fatal COVID-19 is consistent with the hypothesis that anellovirus load may be higher in patients with more severe COVID-19.

Although only a small number of individuals carry eHHV-6 or a high load of anellovirus, positivity appears to have a striking impact on disease risk and clinical phenotype, especially compared with previously identified genetic and environmental factors of these diseases<sup>20–22</sup>. This suggests that the presence of these viruses in an individual's virome is a potentially clinically useful biomarker for these diseases, enabling a personalized medicine approach to prevent disease. In addition to disease risk, eHHV-6 may be an important factor in clinical practice, such as cell therapies<sup>51</sup>. With the expanding availability of clinical WGS, incorporating bioinformatic pipelines that assay the virome may be a cost-effective way to screen for these individuals. The link between endogenous viruses or viral infection and immune-related diseases is a longstanding mystery in basic and clinical science. This study provides part of the missing pieces and outlines a path through which the study of the human virome can be integrated into the genotype–phenotype

association analysis for which human WGS are often collected. Additional clinical and translational studies are needed to better assess the pathologically relevant mechanisms through which eHHV-6 and anellovirus modulate immunity.

### Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-024-02022-z>.

### References

- Hardy, T. A., Blum, S., McCombe, P. A. & Reddel, S. W. Guillain-Barré syndrome: modern theories of etiology. *Curr. Allergy Asthma Rep.* **11**, 197–204 (2011).
- Bjornevik, K. et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* **375**, 296–301 (2022).
- Lanz, T. V. et al. Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GialCAM. *Nature* **603**, 321–327 (2022).
- Miller, S. D. et al. Persistent infection with Theiler's virus leads to CNS autoimmunity via epitope spreading. *Nat. Med.* **3**, 1133–1136 (1997).
- Tokuyama, M. et al. Antibodies against human endogenous retrovirus K102 envelope activate neutrophils in systemic lupus erythematosus. *J. Exp. Med.* **218**, e20191766 (2021).
- Fajgenbaum, D. C. & June, C. H. Cytokine storm. *N. Engl. J. Med.* **383**, 2255–2273 (2020).
- Ikuta, K., Satoh, Y., Hoshikawa, Y. & Sairenji, T. Detection of Epstein-Barr virus in salivas and throat washings in healthy adults and children. *Microbes Infect.* **2**, 115–120 (2000).
- Moustafa, A. et al. The blood DNA virome in 8,000 humans. *PLoS Pathog.* **13**, e1006292 (2017).
- Liu, X. et al. Endogenization and excision of human herpesvirus 6 in human genomes. *PLoS Genet.* **16**, e1008915 (2020).
- Wood, M. L. et al. Variation in human herpesvirus 6B telomeric integration, excision, and transmission between tissues and individuals. *eLife* **10**, e70452 (2021).
- Kojima, S., Kamada, A. J. & Parrish, N. F. Virus-derived variation in diverse human genomes. *PLoS Genet.* **17**, e1009324 (2021).
- Aswad, A. et al. Evolutionary history of endogenous human Herpesvirus 6 reflects human migration out of Africa. *Mol. Biol. Evol.* **38**, 96–107 (2021).

13. Kumata, R., Ito, J., Takahashi, K., Suzuki, T. & Sato, K. A tissue level atlas of the healthy human virome. *BMC Biol.* **18**, 55 (2020).
14. Abbas, A. A. et al. Redondoviridae, a family of small, circular DNA viruses of the human oro-respiratory tract associated with periodontitis and critical illness. *Cell Host Microbe* **25**, 719–729.e4 (2019).
15. Abbas, A. A. et al. The perioperative lung transplant virome: torque teno viruses are elevated in donor lungs and show divergent dynamics in primary graft dysfunction. *Am. J. Transpl.* **17**, 1313–1324 (2017).
16. Maggi, F. et al. TT virus loads and lymphocyte subpopulations in children with acute respiratory diseases. *J. Virol.* **77**, 9081–9083 (2003).
17. Gravel, A. et al. Inherited chromosomally integrated human herpesvirus 6 as a predisposing risk factor for the development of angina pectoris. *Proc. Natl Acad. Sci. USA* **112**, 8058–8063 (2015).
18. Gaccioli, F. et al. Fetal inheritance of chromosomally integrated human herpesvirus 6 predisposes the mother to pre-eclampsia. *Nat. Microbiol.* **5**, 901–908 (2020).
19. Peddu, V. et al. Inherited chromosomally integrated human herpesvirus 6 demonstrates tissue-specific RNA expression in vivo that correlates with an increased antibody immune response. *J. Virol.* **94**, e01418-19 (2019).
20. Sakaue, S. et al. Genetic determinants of risk in autoimmune pulmonary alveolar proteinosis. *Nat. Commun.* **12**, 1032 (2021).
21. Okada, Y. et al. A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus erythematosus in Japanese. *PLoS Genet.* **8**, e1002455 (2012).
22. Ogawa, K. et al. Next-generation sequencing identifies contribution of both class I and II HLA genes on susceptibility of multiple sclerosis in Japanese. *J. Neuroinflammation* **16**, 162 (2019).
23. Bombardier, C., Gladman, D. D., Urowitz, M. B., Caron, D. & Chang, C. H. Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. *Arthritis Rheum.* **35**, 630–640 (1992).
24. Arbuckle, M. R. et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N. Engl. J. Med.* **349**, 1526–1533 (2003).
25. Latzin, P. et al. Anti-GM-CSF antibodies in paediatric pulmonary alveolar proteinosis. *Thorax* **60**, 39–44 (2005).
26. All of Us Research Program Investigators et al. The ‘All of Us’ research program. *N. Engl. J. Med.* **381**, 668–676 (2019).
27. Huang, Y. et al. Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome. *Nucleic Acids Res.* **42**, 315–327 (2014).
28. Mohan, D. et al. PhIP-seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nat. Protoc.* **13**, 1958–1978 (2018).
29. Shrock, E. L., Shrock, C. L. & Elledge, S. J. VirScan: high-throughput profiling of antiviral antibody epitopes. *Bio Protoc.* **12**, e4464 (2022).
30. De Bolle, L., Naesens, L. & De Clercq, E. Update on human herpesvirus 6 biology, clinical features, and therapy. *Clin. Microbiol. Rev.* **18**, 217–245 (2005).
31. Takemoto, K. et al. Time-series analysis comparing the prevalence of antibodies against nine viral species found in umbilical cord blood in Japan. *Jpn. J. Infect. Dis.* **69**, 314–318 (2016).
32. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
33. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).
34. Tanaka-Taya, K. et al. Human herpesvirus 6 (HHV-6) is transmitted from parent to child in an integrated form and characterization of cases with chromosomally integrated HHV-6 DNA. *J. Med. Virol.* **73**, 465–473 (2004).
35. Miura, H. et al. Chromosomally integrated human herpesvirus 6 in the Japanese population. *J. Med. Virol.* **90**, 1636–1642 (2018).
36. Kumata, R., Ito, J. & Sato, K. Inherited chromosomally integrated HHV-6 possibly modulates human gene expression. *Virus Genes* **56**, 386–389 (2020).
37. De Munnik, S. M., Smit, M. J., Leurs, R. & Vischer, H. F. Modulation of cellular signaling by herpesvirus-encoded G protein-coupled receptors. *Front. Pharmacol.* **6**, 40 (2015).
38. Isegawa, Y., Ping, Z., Nakano, K., Sugimoto, N. & Yamanishi, K. Human herpesvirus 6 open reading frame U12 encodes a functional  $\beta$ -chemokine receptor. *J. Virol.* **72**, 6104–6112 (1998).
39. Broccolo, F. et al. Selective reactivation of human herpesvirus 6 in patients with autoimmune connective tissue diseases. *J. Med. Virol.* **85**, 1925–1934 (2013).
40. Worobey, M. Extensive homologous recombination among widely divergent TT viruses. *J. Virol.* **74**, 7666–7670 (2000).
41. Cebriá-Mendoza, M. et al. Deep viral blood metagenomics reveals extensive anellovirus diversity in healthy humans. *Sci. Rep.* **11**, 6921 (2021).
42. Arze, C. A. et al. Global genome analysis reveals a vast and dynamic anellovirus landscape within the human virome. *Cell Host Microbe* **29**, 1305–1315.e6 (2021).
43. De Vlaminck, I. et al. Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell* **155**, 1178–1187 (2013).
44. Shibayama, T. et al. Inverse relationship between the titre of TT virus DNA and the CD4 cell count in patients infected with HIV. *AIDS* **15**, 563–570 (2001).
45. Stoermer, K. A. & Morrison, T. E. Complement and viral pathogenesis. *Virology* **411**, 362–373 (2011).
46. Jung, J.-Y. & Suh, C.-H. Infection in systemic lupus erythematosus, similarities, and differences with lupus flare. *Korean J. Intern. Med.* **32**, 429–438 (2017).
47. Zheng, H. et al. Torque teno virus (SANBAN isolate) ORF2 protein suppresses NF- $\kappa$ B pathways via interaction with I $\kappa$ B kinases. *J. Virol.* **81**, 11917–11924 (2007).
48. Rocchi, J. et al. Torquetenovirus DNA drives proinflammatory cytokines production and secretion by immune cells via toll-like receptor 9. *Virology* **394**, 235–242 (2009).
49. Sospedra, M. et al. Recognition of conserved amino acid motifs of common viruses and its role in autoimmunity. *PLoS Pathog.* **1**, e41 (2005).
50. Seemayer, C. A. et al. Prevalence of TTV DNA and GBV-C RNA in patients with systemic sclerosis, rheumatoid arthritis, and osteoarthritis does not differ from that in healthy blood donors. *Ann. Rheum. Dis.* **60**, 806–809 (2001).
51. Lareau, C. A. et al. Latent human herpesvirus 6 is reactivated in CAR T cells. *Nature* **623**, 608–615 (2023).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's

Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain

permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Noah Sasa<sup>1,2,3,4,218</sup>, Shohei Kojima<sup>5,218</sup>, Rie Koide<sup>5,218</sup>, Takanori Hasegawa<sup>6</sup>, Ho Namkoong<sup>7</sup>, Tomomitsu Hirota<sup>8</sup>, Rei Watanabe<sup>9,10</sup>, Yuumi Nakamura<sup>11,12</sup>, Eri Oguro-Igashira<sup>13,14</sup>, Kotaro Ogawa<sup>1,15</sup>, Tomohiro Yata<sup>1,15</sup>, Kyuto Sonehara<sup>1,3,4</sup>, Kenichi Yamamoto<sup>1,16,17</sup>, Toshihiro Kishikawa<sup>1,2,18</sup>, Saori Sakaue<sup>1,19,20,21</sup>, Ryuya Edahiro<sup>1,13</sup>, Yuya Shirai<sup>1,13,22</sup>, Yuichi Maeda<sup>13,14</sup>, Takuro Nii<sup>13,14</sup>, Shotaro Chubachi<sup>23</sup>, Hiromu Tanaka<sup>23</sup>, Haruka Yabukami<sup>5</sup>, Akari Suzuki<sup>24</sup>, Kimiko Nakajima<sup>25</sup>, Noriko Arase<sup>26</sup>, Takashi Okamoto<sup>27</sup>, Rika Nishikawa<sup>28</sup>, Shinichi Namba<sup>1,3,4</sup>, Tatsuhiko Naito<sup>1,3,4</sup>, Ippei Miyagawa<sup>29</sup>, Hiroaki Tanaka<sup>29</sup>, Masanobu Ueno<sup>29</sup>, Yosuke Ishitsuka<sup>9,10</sup>, Junichi Furuta<sup>10</sup>, Kayo Kunimoto<sup>30</sup>, Ikko Kajihara<sup>31</sup>, Satoshi Fukushima<sup>31</sup>, Hideaki Miyachi<sup>11</sup>, Hiroyuki Matsue<sup>11</sup>, Masahiro Kamata<sup>32</sup>, Mami Momose<sup>33</sup>, Toshinori Bito<sup>28</sup>, Hiroshi Nagai<sup>28</sup>, Tetsuya Ikeda<sup>34</sup>, Tatsuya Horikawa<sup>35</sup>, Atsuko Adachi<sup>36</sup>, Tsukasa Matsubara<sup>37</sup>, Kyoko Ikumi<sup>38</sup>, Emi Nishida<sup>38</sup>, Ikuma Nakagawa<sup>39</sup>, Mayu Yagita-Sakamaki<sup>13,14</sup>, Maiko Yoshimura<sup>40</sup>, Shiro Ohshima<sup>40</sup>, Makoto Kinoshita<sup>15</sup>, Satoru Ito<sup>41</sup>, Toru Arai<sup>42</sup>, Masaki Hirose<sup>42</sup>, Yoshinori Tanino<sup>43</sup>, Takefumi Nikaido<sup>43</sup>, Toshio Ichiwata<sup>44</sup>, Shinya Ohkouchi<sup>45</sup>, Taizou Hirano<sup>46</sup>, Toshinori Takada<sup>47</sup>, Ryushi Tazawa<sup>48</sup>, Konosuke Morimoto<sup>49</sup>, Masahiro Takaki<sup>50</sup>, Satoshi Konno<sup>51</sup>, Masaru Suzuki<sup>51</sup>, Keisuke Tomii<sup>52</sup>, Atsushi Nakagawa<sup>52</sup>, Tomohiro Handa<sup>53</sup>, Kiminobu Tanizawa<sup>54</sup>, Haruyuki Ishii<sup>55</sup>, Manabu Ishida<sup>55</sup>, Toshiyuki Kato<sup>56</sup>, Naoya Takeda<sup>56</sup>, Koshi Yokomura<sup>57</sup>, Takashi Matsui<sup>57</sup>, Akifumi Uchida<sup>58</sup>, Hiromasa Inoue<sup>58</sup>, Kazuyoshi Imaizumi<sup>59</sup>, Yasuhiro Goto<sup>59</sup>, Hiroshi Kida<sup>13,60</sup>, Tomoyuki Fujisawa<sup>61</sup>, Takafumi Suda<sup>61</sup>, Takashi Yamada<sup>62</sup>, Yasuomi Satake<sup>62</sup>, Hidenori Iбата<sup>63</sup>, Mika Saigusa<sup>64</sup>, Toshihiro Shirai<sup>64</sup>, Nobuyuki Hizawa<sup>65</sup>, Koh Nakata<sup>66</sup>, Japan COVID-19 Task Force\*, Shinichi Imafuku<sup>67</sup>, Yayoi Tada<sup>32</sup>, Yoshihide Asano<sup>68,69</sup>, Shinichi Sato<sup>69</sup>, Chikako Nishigori<sup>28</sup>, Masatoshi Jinnin<sup>30</sup>, Hironobu Ihn<sup>31</sup>, Akihiko Asahina<sup>33</sup>, Hidehisa Saeki<sup>70</sup>, Tatsuyoshi Kawamura<sup>27</sup>, Shinji Shimada<sup>27</sup>, Ichiro Katayama<sup>26,71</sup>, Hannah M. Poisner<sup>72</sup>, Taralynn M. Mack<sup>72</sup>, Alexander G. Bick<sup>72</sup>, Koichiro Higasa<sup>73,74</sup>, Tatsusada Okuno<sup>15</sup>, Hideki Mochizuki<sup>15</sup>, Makoto Ishii<sup>23,75</sup>, Ryuji Koike<sup>76</sup>, Akinori Kimura<sup>77</sup>, Emiko Noguchi<sup>78</sup>, Shigetoshi Sano<sup>25</sup>, Hidenori Inohara<sup>2</sup>, Manabu Fujimoto<sup>10,26</sup>, Yoshikazu Inoue<sup>42</sup>, Etsuro Yamaguchi<sup>41</sup>, Seishi Ogawa<sup>79,80,81</sup>, Takanori Kanai<sup>82</sup>, Akimichi Morita<sup>38</sup>, Fumihiko Matsuda<sup>73</sup>, Mayumi Tamari<sup>8</sup>, Atsushi Kumanogoh<sup>13,83</sup>, Yoshiya Tanaka<sup>29</sup>, Koichiro Ohmura<sup>84</sup>, Koichi Fukunaga<sup>23</sup>, Seiya Imoto<sup>85</sup>, Satoru Miyano<sup>6</sup>, Nicholas F. Parrish<sup>5</sup> ✉ & Yukinori Okada<sup>1,3,4,22,86</sup> ✉

<sup>1</sup>Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. <sup>2</sup>Department of Otorhinolaryngology—Head and Neck Surgery, Osaka University Graduate School of Medicine, Suita, Japan. <sup>3</sup>Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. <sup>4</sup>Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. <sup>5</sup>Genome Immunobiology RIKEN Hakubi Research Team, RIKEN Center for Integrative Medical Sciences and RIKEN Cluster for Pioneering Research, Yokohama, Japan. <sup>6</sup>M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan. <sup>7</sup>Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan. <sup>8</sup>Division of Molecular Genetics, Research Center for Medical Science, The Jikei University School of Medicine, Tokyo, Japan. <sup>9</sup>Department of Integrative Medicine for Allergic and Immunological Diseases, Osaka University Graduate School of Medicine, Suita, Japan. <sup>10</sup>Department of Dermatology, University of Tsukuba, Tsukuba, Japan. <sup>11</sup>Department of Dermatology, Chiba University Graduate School of Medicine, Chiba, Japan. <sup>12</sup>Cutaneous Immunology, Immunology Frontier Research Center, Osaka University, Suita, Japan. <sup>13</sup>Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, Suita, Japan. <sup>14</sup>Laboratory of Immune Regulation, Department of Microbiology and Immunology, Osaka University Graduate School of Medicine, Suita, Japan. <sup>15</sup>Department of Neurology, Osaka University Graduate School of Medicine, Suita, Japan. <sup>16</sup>Department of Pediatrics, Osaka University Graduate School of Medicine, Suita, Japan. <sup>17</sup>Division of Health Science, Osaka University Graduate School of Medicine, Suita, Japan. <sup>18</sup>Department of Head and Neck Surgery, Aichi Cancer Center Hospital, Nagoya, Japan. <sup>19</sup>Center for Data Sciences, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>20</sup>Divisions of Genetics and Rheumatology, Brigham and Women's Hospital and Department of Medicine, Harvard Medical School, Boston, MA, USA. <sup>21</sup>Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>22</sup>Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan. <sup>23</sup>Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo, Japan. <sup>24</sup>Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. <sup>25</sup>Department of Dermatology, Kochi Medical School, Kochi University, Nankoku, Japan. <sup>26</sup>Department of Dermatology, Osaka University Graduate School of Medicine, Suita, Japan. <sup>27</sup>Department of Dermatology, Faculty of Medicine, University of Yamanashi, Yamanashi, Japan. <sup>28</sup>Division of Dermatology, Department of Internal Related, Faculty of Medicine, Kobe University Graduate School of Medicine, Kobe, Japan. <sup>29</sup>The First Department of Internal Medicine, University of Occupational and Environmental Health, Kitakyushu, Japan. <sup>30</sup>Department of Dermatology, Wakayama Medical University Graduate School of Medicine, Wakayama, Japan. <sup>31</sup>Department of Dermatology and Plastic Surgery, Kumamoto University, Kumamoto, Japan. <sup>32</sup>Department of Dermatology, Teikyo University School of Medicine, Tokyo, Japan. <sup>33</sup>Department of Dermatology, The Jikei University School of Medicine, Tokyo, Japan. <sup>34</sup>Department of Dermatology, Kobe City Hospital Organization Kobe Medical Center West Hospital, Kobe, Japan. <sup>35</sup>Department of Dermatology, Nishikobe Medical Center, Kobe, Japan. <sup>36</sup>Department of Dermatology, Hyogo Prefectural Kakogawa Medical Center, Kakogawa, Japan. <sup>37</sup>Matsubara Mayflower Hospital, Kato, Japan. <sup>38</sup>Department of Geriatric and Environmental Dermatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan. <sup>39</sup>Hokkaido Medical Center for Rheumatic Diseases, Sapporo, Japan. <sup>40</sup>Department of Rheumatology, NHO Osaka Minami Medical Center, Kawachinagano, Japan. <sup>41</sup>Division of Respiratory Medicine and Allergology,

Department of Internal Medicine, School of Medicine, Aichi Medical University, Aichi, Japan. <sup>42</sup>Clinical Research Center, NHO Kinki Chuo Chest Medical Center, Sakai, Japan. <sup>43</sup>Department of Pulmonary Medicine, Fukushima Medical University, Fukushima, Japan. <sup>44</sup>Department Respiratory Medicine, Tokyo Medical University, Tokyo, Japan. <sup>45</sup>Department of Occupational Health, Tohoku University Graduate School of Medicine, Sendai, Japan. <sup>46</sup>Hirosaki Medical Center, Hirosaki, Japan. <sup>47</sup>Uonuma Institute of Community Medicine, Niigata University Medical and Dental Hospital, Niigata, Japan. <sup>48</sup>Student Support and Health Administration Organization, Tokyo Medical and Dental University, Tokyo, Japan. <sup>49</sup>Department of Respiratory Infections, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan. <sup>50</sup>Department of Infectious Diseases, Nagasaki University Hospital, Nagasaki University, Nagasaki, Japan. <sup>51</sup>Department of Respiratory Medicine, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo, Japan. <sup>52</sup>Department of Respiratory Medicine, Kobe City Medical Center General Hospital, Kobe, Japan. <sup>53</sup>Department of Advanced Medicine for Respiratory Failure, Graduate School of Medicine, Kyoto University, Kyoto, Japan. <sup>54</sup>Department of Respiratory Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan. <sup>55</sup>Department of Respiratory Medicine, Kyorin University Faculty of Medicine, Mitaka, Japan. <sup>56</sup>Department of Respiratory Medicine and Allergology, Kariya Toyota General Hospital, Kariya, Japan. <sup>57</sup>Department of Respiratory Medicine, Respiratory Disease Center, Seirei Mikatahara General Hospital, Hamamatsu, Japan. <sup>58</sup>Department of Pulmonary Medicine, Graduate School of Medical & Dental Sciences, Kagoshima University, Kagoshima, Japan. <sup>59</sup>Department of Respiratory Medicine, Fujita Health University School of Medicine, Aichi, Japan. <sup>60</sup>Department of Respiratory Medicine, National Hospital Organization Osaka Toneyama Medical Center, Toyonaka, Japan. <sup>61</sup>Second Division, Department of Internal Medicine, Hamamatsu University School of Medicine, Hamamatsu, Japan. <sup>62</sup>Department of Respiratory Medicine, Shizuoka City Shizuoka Hospital, Shizuoka, Japan. <sup>63</sup>Department of Respiratory Medicine, National Hospital Organization Mie Chuo Medical Center, Tsu, Japan. <sup>64</sup>Department of Respiratory Medicine, Shizuoka General Hospital, Shizuoka, Japan. <sup>65</sup>Department of Pulmonary Medicine, Institute of Medicine, University of Tsukuba, Tsukuba, Japan. <sup>66</sup>Division of Advanced Medical Development, Niigata University Medical and Dental Hospital, Niigata, Japan. <sup>67</sup>Department of Dermatology, Faculty of Medicine, Fukuoka University, Fukuoka, Japan. <sup>68</sup>Department of Dermatology, Tohoku University Graduate School of Medicine, Sendai, Japan. <sup>69</sup>Department of Dermatology, University of Tokyo Graduate School of Medicine, Tokyo, Japan. <sup>70</sup>Department of Dermatology, Nippon Medical School, Tokyo, Japan. <sup>71</sup>Department of Pigmentation Research and Therapeutics, Osaka City University Graduate School of Medicine, Osaka, Japan. <sup>72</sup>Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN, USA. <sup>73</sup>Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan. <sup>74</sup>Department of Genome Analysis, Institute of Biomedical Science, Kansai Medical University, Hirakata, Japan. <sup>75</sup>Department of Respiratory Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan. <sup>76</sup>Health Science Research and Development Center (HeRD), Tokyo Medical and Dental University, Tokyo, Japan. <sup>77</sup>Institute of Research, Tokyo Medical and Dental University, Tokyo, Japan. <sup>78</sup>Department of Medical Genetics, University of Tsukuba, Tsukuba, Japan. <sup>79</sup>Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan. <sup>80</sup>Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan. <sup>81</sup>Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden. <sup>82</sup>Division of Gastroenterology and Hepatology, Department of Medicine, Keio University School of Medicine, Tokyo, Japan. <sup>83</sup>Department of Immunopathology, Immunology Frontier Research Center, Osaka University, Suita, Japan. <sup>84</sup>Department of Rheumatology, Kobe City Medical Center General Hospital, Kobe, Japan. <sup>85</sup>Division of Health Medical Intelligence, Human Genome Center, the Institute of Medical Science, the University of Tokyo, Tokyo, Japan. <sup>86</sup>Premium Research Institute for Human Metaverse Medicine (WPI-PRiMe), Osaka University, Suita, Japan. <sup>218</sup>These authors contributed equally: Noah Sasa, Shohei Kojima, Rie Koide. \*A list of authors and their affiliations appears at the end of the paper. ✉e-mail: [nicholas.parrish@riken.jp](mailto:nicholas.parrish@riken.jp); [yuki-okada@m.u-tokyo.ac.jp](mailto:yuki-okada@m.u-tokyo.ac.jp)

## Japan COVID-19 Task Force

**Qingbo S. Wang<sup>1,22</sup>, Ryuya Edahiro<sup>1,13</sup>, Ho Namkoong<sup>7</sup>, Takanori Hasegawa<sup>6</sup>, Yuya Shirai<sup>1,13,22</sup>, Kyuto Sonehara<sup>1,3,4</sup>, Hiromu Tanaka<sup>23</sup>, Ho Lee<sup>23</sup>, Ryunosuke Saiki<sup>79</sup>, Takayoshi Hyugaji<sup>85</sup>, Eigo Shimizu<sup>85</sup>, Kotoe Katayama<sup>85</sup>, Masahiro Kanai<sup>87</sup>, Tatsuhiko Naito<sup>1,3,4</sup>, Noah Sasa<sup>1,2,3,4,218</sup>, Kenichi Yamamoto<sup>1,16,17</sup>, Yasuhiro Kato<sup>13,83</sup>, Takayoshi Morita<sup>13,83</sup>, Kazuhisa Takahashi<sup>88</sup>, Norihiro Harada<sup>88</sup>, Toshio Naito<sup>89</sup>, Makoto Hiki<sup>90,91</sup>, Yasushi Matsushita<sup>92</sup>, Haruhi Takagi<sup>88</sup>, Masako Ichikawa<sup>88</sup>, Ai Nakamura<sup>88</sup>, Sonoko Harada<sup>88,93</sup>, Yuuki Sandhu<sup>88</sup>, Hiroki Kabata<sup>23</sup>, Katsunori Masaki<sup>23</sup>, Hirofumi Kamata<sup>23</sup>, Shinnosuke Ikemura<sup>23</sup>, Shotaro Chubachi<sup>23</sup>, Satoshi Okamori<sup>23</sup>, Hideki Terai<sup>23</sup>, Atsuhito Morita<sup>23</sup>, Takanori Asakura<sup>23</sup>, Junichi Sasaki<sup>94</sup>, Hiroshi Morisaki<sup>95</sup>, Yoshifumi Uwamino<sup>96</sup>, Kosaku Nanki<sup>82</sup>, Sho Uchida<sup>7</sup>, Shunsuke Uno<sup>7</sup>, Tomoyasu Nishimura<sup>7,97</sup>, Takashri Ishiguro<sup>98</sup>, Taisuke Isono<sup>98</sup>, Shun Shibata<sup>98</sup>, Yuma Matsui<sup>98</sup>, Chiaki Hosoda<sup>98</sup>, Kenji Takano<sup>98</sup>, Takashi Nishida<sup>98</sup>, Yoichi Kobayashi<sup>98</sup>, Yotaro Takaku<sup>98</sup>, Noboru Takayanagi<sup>98</sup>, Soichiro Ueda<sup>99</sup>, Ai Tada<sup>99</sup>, Masayoshi Miyawaki<sup>99</sup>, Masaomi Yamamoto<sup>99</sup>, Eriko Yoshida<sup>99</sup>, Reina Hayashi<sup>99</sup>, Tomoki Nagasaka<sup>99</sup>, Sawako Arai<sup>99</sup>, Yutaro Kaneko<sup>99</sup>, Kana Sasaki<sup>99</sup>, Etsuko Tagaya<sup>100</sup>, Masatoshi Kawana<sup>101</sup>, Ken Arimura<sup>100</sup>, Kunihiko Takahashi<sup>6</sup>, Tatsuhiko Anzai<sup>6</sup>, Satoshi Ito<sup>6</sup>, Akifumi Endo<sup>102</sup>, Yuji Uchimura<sup>103</sup>, Yasunari Miyazaki<sup>104</sup>, Takayuki Honda<sup>104</sup>, Tomoya Tateishi<sup>104</sup>, Shuji Tohda<sup>105</sup>, Naoya Ichimura<sup>105</sup>, Kazunari Sonobe<sup>105</sup>, Chihito Tani Sassa<sup>105</sup>, Jun Nakajima<sup>105</sup>, Yasushi Nakano<sup>106</sup>, Yukiko Nakajima<sup>106</sup>, Ryusuke Anan<sup>106</sup>, Ryosuke Arai<sup>106</sup>, Yuko Kurihara<sup>106</sup>, Yuko Harada<sup>106</sup>, Kazumi Nishio<sup>106</sup>, Tetsuya Ueda<sup>107</sup>, Masanori Azuma<sup>107</sup>, Ryuichi Saito<sup>107</sup>, Toshikatsu Sado<sup>107</sup>, Yoshimune Miyazaki<sup>107</sup>, Ryuichi Sato<sup>107</sup>, Yuki Haruta<sup>107</sup>, Tadao Nagasaki<sup>107</sup>, Yoshinori Yasui<sup>108</sup>, Yoshinori Hasegawa<sup>107</sup>, Yoshikazu Mutoh<sup>109</sup>, Tomoki Kimura<sup>110</sup>, Tomonori Sato<sup>110</sup>, Reoto Takei<sup>110</sup>, Satoshi Hagimoto<sup>110</sup>, Yoichiro Noguchi<sup>110</sup>, Yasuhiko Yamano<sup>110</sup>, Hajime Sasano<sup>110</sup>, Sho Ota<sup>110</sup>, Yasushi Nakamori<sup>111</sup>, Kazuhisa Yoshiya<sup>111</sup>, Fukuki Saito<sup>111</sup>, Tomoyuki Yoshihara<sup>111</sup>, Daiki Wada<sup>111</sup>, Hiromu Iwamura<sup>111</sup>, Syuji Kanayama<sup>111</sup>, Shuhei Maruyama<sup>111</sup>, Takashi Yoshiyama<sup>112</sup>, Ken Ohta<sup>112</sup>, Hiroyuki Kokuto<sup>112</sup>, Hideo Ogata<sup>112</sup>, Yoshiaki Tanaka<sup>112</sup>, Kenichi Arakawa<sup>112</sup>, Masafumi Shimoda<sup>112</sup>, Takeshi Osawa<sup>112</sup>, Hiroki Tateno<sup>113</sup>, Isano Hase<sup>113</sup>, Shuichi Yoshida<sup>113</sup>, Shoji Suzuki<sup>113</sup>, Miki Kawada<sup>114</sup>, Hirohisa Horinouchi<sup>115</sup>, Fumitake Saito<sup>116</sup>, Keiko Mitamura<sup>117</sup>, Masao Hagihara<sup>118</sup>, Junichi Ochi<sup>116</sup>, Tomoyuki Uchida<sup>118</sup>, Rie Baba<sup>119</sup>, Daisuke Arai<sup>119</sup>, Takayuki Ogura<sup>119</sup>, Hidenori Takahashi<sup>119</sup>, Shigehiro Hagiwara<sup>119</sup>, Genta Nagao<sup>119</sup>, Shunichiro Konishi<sup>119</sup>, Ichiro Nakachi<sup>119</sup>, Koji Murakami<sup>120</sup>, Mitsuhiro Yamada<sup>120</sup>, Hisatoshi Sugiura<sup>120</sup>,**

Hirohito Sano<sup>120</sup>, Shuichiro Matsumoto<sup>120</sup>, Nozomu Kimura<sup>120</sup>, Yoshinao Ono<sup>120</sup>, Hiroaki Baba<sup>121</sup>, Yusuke Suzuki<sup>122</sup>, Sohei Nakayama<sup>122</sup>, Keita Masuzawa<sup>122</sup>, Shinichi Namba<sup>1,3,4</sup>, Takayuki Shiroyama<sup>13</sup>, Yoshimi Noda<sup>13</sup>, Takayuki Niitsu<sup>13</sup>, Yuichi Adachi<sup>13</sup>, Takatoshi Enomoto<sup>13</sup>, Saori Amiya<sup>13</sup>, Reina Hara<sup>13</sup>, Yuta Yamaguchi<sup>13,83</sup>, Teruaki Murakami<sup>13,83</sup>, Tomoki Kuge<sup>13</sup>, Kinnosuke Matsumoto<sup>13</sup>, Yuji Yamamoto<sup>13</sup>, Makoto Yamamoto<sup>13</sup>, Midori Yoneda<sup>13</sup>, Kazunori Tomono<sup>123</sup>, Kazuto Kato<sup>124</sup>, Haruhiko Hirata<sup>13</sup>, Yoshito Takeda<sup>13</sup>, Hidesumi Koh<sup>125</sup>, Tadashi Manabe<sup>125</sup>, Yohei Funatsu<sup>125</sup>, Fumimaro Ito<sup>125</sup>, Takahiro Fukui<sup>125</sup>, Keisuke Shinozuka<sup>125</sup>, Sumiko Kohashi<sup>125</sup>, Masatoshi Miyazaki<sup>125</sup>, Tomohisa Shoko<sup>126</sup>, Mitsuaki Kojima<sup>126</sup>, Tomohiro Adachi<sup>126</sup>, Motonao Ishikawa<sup>127</sup>, Kenichiro Takahashi<sup>128</sup>, Takashi Inoue<sup>129</sup>, Toshiyuki Hirano<sup>129</sup>, Keigo Kobayashi<sup>129</sup>, Hatsuyo Takaoka<sup>126</sup>, Kazuyoshi Watanabe<sup>130</sup>, Naoki Miyazawa<sup>131</sup>, Yasuhiro Kimura<sup>131</sup>, Reiko Sado<sup>131</sup>, Hideyasu Sugimoto<sup>131</sup>, Akane Kamiya<sup>132</sup>, Naota Kuwahara<sup>133</sup>, Akiko Fujiwara<sup>133</sup>, Tomohiro Matsunaga<sup>133</sup>, Yoko Sato<sup>133</sup>, Takenori Okada<sup>133</sup>, Yoshihiro Hirai<sup>134</sup>, Hidetoshi Kawashima<sup>134</sup>, Atsuya Narita<sup>134</sup>, Kazuki Niwa<sup>135</sup>, Yoshiyuki Sekikawa<sup>135</sup>, Koichi Nishi<sup>136</sup>, Masaru Nishitsuji<sup>136</sup>, Mayuko Tani<sup>136</sup>, Junya Suzuki<sup>136</sup>, Hiroki Nakatsumi<sup>136</sup>, Takashi Ogura<sup>137</sup>, Hideya Kitamura<sup>137</sup>, Eri Hagiwara<sup>137</sup>, Kota Murohashi<sup>137</sup>, Hiroko Okabayashi<sup>137</sup>, Takao Mochimaru<sup>138,139</sup>, Shigenari Nukaga<sup>138</sup>, Ryosuke Satomi<sup>138</sup>, Yoshitaka Oyamada<sup>138,139</sup>, Nobuaki Mori<sup>140</sup>, Tomoya Baba<sup>141</sup>, Yasutaka Fukui<sup>141</sup>, Mitsuru Odate<sup>141</sup>, Shuko Mashimo<sup>141</sup>, Yasushi Makino<sup>141</sup>, Kazuma Yagi<sup>142</sup>, Mizuha Hashiguchi<sup>142</sup>, Junko Kagyo<sup>142</sup>, Tetsuya Shiomi<sup>142</sup>, Satoshi Fuke<sup>143</sup>, Hiroshi Saito<sup>143</sup>, Tomoya Tsuchida<sup>144</sup>, Shigeki Fujitani<sup>145</sup>, Mumon Takita<sup>145</sup>, Daiki Morikawa<sup>145</sup>, Toru Yoshida<sup>145</sup>, Takehiro Izumo<sup>146</sup>, Minoru Inomata<sup>146</sup>, Naoyuki Kuse<sup>146</sup>, Nobuyasu Awano<sup>146</sup>, Mari Tone<sup>146</sup>, Akihiro Ito<sup>147</sup>, Yoshihiko Nakamura<sup>148</sup>, Kota Hoshino<sup>148</sup>, Junichi Maruyama<sup>148</sup>, Hiroyasu Ishikura<sup>148</sup>, Tooru Takata<sup>149</sup>, Toshio Odani<sup>150</sup>, Masaru Amishima<sup>151</sup>, Takeshi Hattori<sup>151</sup>, Yasuo Shichinohe<sup>152</sup>, Takashi Kagaya<sup>153</sup>, Toshiyuki Kita<sup>153</sup>, Kazuhide Ohta<sup>153</sup>, Satoru Sakagami<sup>153</sup>, Kiyoshi Koshida<sup>153</sup>, Kentaro Hayashi<sup>154</sup>, Tetsuo Shimizu<sup>154</sup>, Yutaka Kozu<sup>154</sup>, Hisato Hiranuma<sup>154</sup>, Yasuhiro Gon<sup>154</sup>, Namiki Izumi<sup>155</sup>, Kaoru Nagata<sup>155</sup>, Ken Ueda<sup>155</sup>, Reiko Taki<sup>155</sup>, Satoko Hanada<sup>155</sup>, Kodai Kawamura<sup>156</sup>, Kazuya Ichikado<sup>156</sup>, Kenta Nishiyama<sup>156</sup>, Hiroyuki Muranaka<sup>156</sup>, Kazunori Nakamura<sup>156</sup>, Naozumi Hashimoto<sup>75</sup>, Keiko Wakahara<sup>75</sup>, Sakamoto Koji<sup>75</sup>, Norihito Omote<sup>75</sup>, Akira Ando<sup>75</sup>, Nobuhiro Kodama<sup>157</sup>, Yasunari Kaneyama<sup>157</sup>, Shunsuke Maeda<sup>157</sup>, Takashige Kuraki<sup>158</sup>, Takemasa Matsumoto<sup>158</sup>, Koutaro Yokote<sup>159</sup>, Taka-Aki Nakada<sup>160</sup>, Ryuzo Abe<sup>160</sup>, Taku Oshima<sup>160</sup>, Tadanaga Shimada<sup>160</sup>, Masahiro Harada<sup>161</sup>, Takeshi Takahashi<sup>161</sup>, Hiroshi Ono<sup>161</sup>, Toshihiro Sakurai<sup>161</sup>, Takayuki Shibusawa<sup>161</sup>, Yoshifumi Kimizuka<sup>162</sup>, Akihiko Kawana<sup>162</sup>, Tomoya Sano<sup>162</sup>, Chie Watanabe<sup>162</sup>, Ryohei Suematsu<sup>162</sup>, Hisako Sageshima<sup>163</sup>, Ayumi Yoshifuji<sup>164</sup>, Kazuto Ito<sup>164</sup>, Saeko Takahashi<sup>165</sup>, Kota Ishioka<sup>165</sup>, Morio Nakamura<sup>165</sup>, Makoto Masuda<sup>166</sup>, Aya Wakabayashi<sup>166</sup>, Hiroki Watanabe<sup>166</sup>, Suguru Ueda<sup>166</sup>, Masanori Nishikawa<sup>166</sup>, Yusuke Chihara<sup>167</sup>, Mayumi Takeuchi<sup>167</sup>, Keisuke Onoi<sup>167</sup>, Jun Shinozuka<sup>167</sup>, Atsushi Sueyoshi<sup>167</sup>, Yoji Nagasaki<sup>168</sup>, Masaki Okamoto<sup>169,170</sup>, Sayoko Ishihara<sup>171</sup>, Masatoshi Shimo<sup>171</sup>, Yoshihisa Tokunaga<sup>169,170</sup>, Yu Kusaka<sup>172</sup>, Takehiko Ohba<sup>172</sup>, Susumu Isogai<sup>172</sup>, Aki Ogawa<sup>172</sup>, Takuya Inoue<sup>172</sup>, Satoru Fukuyama<sup>173</sup>, Yoshihiro Eriguchi<sup>174</sup>, Akiko Yonekawa<sup>174</sup>, Keiko Kan-O<sup>173</sup>, Koichiro Matsumoto<sup>173</sup>, Kensuke Kanaoka<sup>175</sup>, Shoichi Ihara<sup>175</sup>, Kiyoshi Komuta<sup>175</sup>, Yoshiaki Inoue<sup>176</sup>, Shigeru Chiba<sup>177</sup>, Kunihiro Yamagata<sup>178</sup>, Yuji Hiramatsu<sup>179</sup>, Hirayasu Kai<sup>178</sup>, Koichiro Asano<sup>180</sup>, Tsuyoshi Oguma<sup>180</sup>, Yoko Ito<sup>180</sup>, Satoru Hashimoto<sup>181</sup>, Masaki Yamasaki<sup>181</sup>, Yu Kasamatsu<sup>182</sup>, Yuko Komase<sup>183</sup>, Naoya Hida<sup>183</sup>, Takahiro Tsuburai<sup>183</sup>, Baku Oyama<sup>183</sup>, Minoru Takada<sup>184</sup>, Hidenori Kanda<sup>184</sup>, Yuichiro Kitagawa<sup>185</sup>, Tetsuya Fukuta<sup>185</sup>, Takahito Miyake<sup>185</sup>, Shozo Yoshida<sup>185</sup>, Shinji Ogura<sup>185</sup>, Shinji Abe<sup>44</sup>, Yuta Kono<sup>44</sup>, Yuki Togashi<sup>44</sup>, Hiroyuki Takoi<sup>44</sup>, Ryota Kikuchi<sup>44</sup>, Shinichi Ogawa<sup>186</sup>, Tomouki Ogata<sup>186</sup>, Shoichiro Ishihara<sup>186</sup>, Arihiko Kanehiro<sup>187,188</sup>, Shinji Ozaki<sup>187</sup>, Yasuko Fuchimoto<sup>187</sup>, Sae Wada<sup>187</sup>, Nobukazu Fujimoto<sup>187</sup>, Kei Nishiyama<sup>189</sup>, Mariko Terashima<sup>190</sup>, Satoru Beppu<sup>190</sup>, Kosuke Yoshida<sup>190</sup>, Osamu Narumoto<sup>191</sup>, Hideaki Nagai<sup>191</sup>, Nobuharu Ooshima<sup>191</sup>, Mitsuru Motegi<sup>192</sup>, Akira Umeda<sup>193</sup>, Kazuya Miyagawa<sup>194</sup>, Hisato Shimada<sup>195</sup>, Mayu Endo<sup>196</sup>, Yoshiyuki Ohira<sup>193</sup>, Masafumi Watanabe<sup>197</sup>, Sumito Inoue<sup>197</sup>, Akira Igarashi<sup>197</sup>, Masamichi Sato<sup>197</sup>, Hironori Sagara<sup>198</sup>, Akihiko Tanaka<sup>198</sup>, Shin Ohta<sup>198</sup>, Tomoyuki Kimura<sup>198</sup>, Yoko Shibata<sup>43</sup>, Yoshinori Tanino<sup>43</sup>, Takefumi Nikaido<sup>43</sup>, Hiroyuki Minemura<sup>43</sup>, Yuki Sato<sup>43</sup>, Yuichiro Yamada<sup>199</sup>, Takuya Hashino<sup>199</sup>, Masato Shinoki<sup>199</sup>, Hajime Iwagoe<sup>200</sup>, Hiroshi Takahashi<sup>201</sup>, Kazuhiko Fujii<sup>201</sup>, Hiroto Kishi<sup>201</sup>, Masayuki Kanai<sup>202</sup>, Tomonori Imamura<sup>202</sup>, Tatsuya Yamashita<sup>202</sup>, Masakiyo Yatomi<sup>203</sup>, Toshitaka Maeno<sup>203</sup>, Shinichi Hayashi<sup>204</sup>, Mai Takahashi<sup>204</sup>, Mizuki Kuramochi<sup>204</sup>, Isamu Kamimaki<sup>204</sup>, Yoshiteru Tominaga<sup>204</sup>, Tomoo Ishii<sup>205</sup>, Mitsuyoshi Utsugi<sup>206</sup>, Akihiro Ono<sup>206</sup>, Toru Tanaka<sup>207</sup>, Takeru Kashiwada<sup>207</sup>, Kazue Fujita<sup>207</sup>, Yoshinobu Saito<sup>207</sup>, Masahiro Seike<sup>207</sup>, Hiroko Watanabe<sup>208</sup>, Hiroto Matsuse<sup>209</sup>, Norio Kodaka<sup>209</sup>, Chihiro Nakano<sup>209</sup>, Takeshi Oshio<sup>209</sup>, Takatomo Hirouchi<sup>209</sup>, Shohei Makino<sup>210</sup>, Moritoki Egi<sup>210</sup>, Yosuke Omae<sup>211</sup>, Yasuhiro Nannya<sup>79</sup>, Takafumi Ueno<sup>212</sup>, Tomomi Takano<sup>213</sup>, Kazuhiko Katayama<sup>214</sup>, Masumi Ai<sup>215</sup>, Atsushi Kumanogoh<sup>13,83</sup>, Toshiro Sato<sup>216</sup>, Naoki Hasegawa<sup>7</sup>, Katsushi Tokunaga<sup>211</sup>, Makoto Ishii<sup>23,75</sup>, Ryuji Koike<sup>76</sup>, Yuko Kitagawa<sup>217</sup>, Akinori Kimura<sup>77</sup>, Seiya Imoto<sup>85</sup>, Satoru Miyano<sup>6</sup>, Seishi Ogawa<sup>79,80,81</sup>, Takanori Kanai<sup>82</sup>, Koichi Fukunaga<sup>23</sup> & Yukinori Okada<sup>1,3,4,22,86</sup>

<sup>87</sup>Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. <sup>88</sup>Department of Respiratory Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan. <sup>89</sup>Department of General Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan. <sup>90</sup>Department of Emergency and Disaster Medicine, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan. <sup>91</sup>Department of Cardiovascular Biology and Medicine, Juntendo University Faculty of Medicine and Graduate School of

Medicine, Tokyo, Japan. <sup>92</sup>Department of Internal Medicine and Rheumatology, Juntendo University Faculty of Medicine and Graduate School of Medicine, Tokyo, Japan. <sup>93</sup>Atopy (Allergy) Research Center, Juntendo University Graduate School of Medicine, Tokyo, Japan. <sup>94</sup>Department of Emergency and Critical Care Medicine, Keio University School of Medicine, Tokyo, Japan. <sup>95</sup>Department of Anesthesiology, Keio University School of Medicine, Tokyo, Japan. <sup>96</sup>Department of Laboratory Medicine, Keio University School of Medicine, Tokyo, Japan. <sup>97</sup>Keio University Health Center, Tokyo, Japan. <sup>98</sup>Department of Respiratory Medicine, Saitama Cardiovascular and Respiratory Center, Kumagaya, Japan. <sup>99</sup>JCHO (Japan Community Health care Organization) Saitama Medical Center, Internal Medicine, Saitama, Japan. <sup>100</sup>Department of Respiratory Medicine, Tokyo Women's Medical University, Tokyo, Japan. <sup>101</sup>Department of General Medicine, Tokyo Women's Medical University, Tokyo, Japan. <sup>102</sup>Clinical Research Center, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan. <sup>103</sup>Department of Medical Informatics, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan. <sup>104</sup>Respiratory Medicine, Tokyo Medical and Dental University, Tokyo, Japan. <sup>105</sup>Clinical Laboratory, Tokyo Medical and Dental University Hospital of Medicine, Tokyo, Japan. <sup>106</sup>Kawasaki Municipal Ida Hospital, Department of Internal Medicine, Kawasaki, Japan. <sup>107</sup>Department of Respiratory Medicine, Osaka Saiseikai Nakatsu Hospital, Osaka, Japan. <sup>108</sup>Department of Infection Control, Osaka Saiseikai Nakatsu Hospital, Osaka, Japan. <sup>109</sup>Department of Infectious Diseases, Tosei General Hospital, Seto, Japan. <sup>110</sup>Department of Respiratory, Allergic Diseases Internal Medicine, Tosei General Hospital, Seto, Japan. <sup>111</sup>Department of Emergency and Critical Care Medicine, Kansai Medical University General Medical Center, Moriguchi, Japan. <sup>112</sup>Japan Anti-Tuberculosis Association (JATA) Fukujji Hospital, Kiyose, Japan. <sup>113</sup>Department of Pulmonary Medicine, Saitama City Hospital, Saitama, Japan. <sup>114</sup>Department of Infectious Diseases, Saitama City Hospital, Saitama, Japan. <sup>115</sup>Department of General Thoracic Surgery, Saitama City Hospital, Saitama, Japan. <sup>116</sup>Department of Pulmonary Medicine, Eiju General Hospital, Tokyo, Japan. <sup>117</sup>Division of Infection Control, Eiju General Hospital, Tokyo, Japan. <sup>118</sup>Department of Hematology, Eiju General Hospital, Tokyo, Japan. <sup>119</sup>Saiseikai Utsunomiya Hospital, Utsunomiya, Japan. <sup>120</sup>Department of Respiratory Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan. <sup>121</sup>Department of Infectious Diseases, Tohoku University Graduate School of Medicine, Sendai, Japan. <sup>122</sup>Department of Respiratory Medicine, Kitasato University Kitasato Institute Hospital, Tokyo, Japan. <sup>123</sup>Division of Infection Control and Prevention, Osaka University Hospital, Suita, Japan. <sup>124</sup>Department of Biomedical Ethics and Public Policy, Osaka University Graduate School of Medicine, Suita, Japan. <sup>125</sup>Tachikawa Hospital, Tachikawa, Japan. <sup>126</sup>Department of Emergency and Critical Care Medicine, Tokyo Women's Medical University Medical Center East, Tokyo, Japan. <sup>127</sup>Department of Medicine, Tokyo Women's Medical University Medical Center East, Tokyo, Japan. <sup>128</sup>Department of Pediatrics, Tokyo Women's Medical University Medical Center East, Tokyo, Japan. <sup>129</sup>Internal Medicine, Sano Kosei General Hospital, Sano, Japan. <sup>130</sup>Japan Community Health Care Organization Kanazawa Hospital, Kanazawa, Japan. <sup>131</sup>Department of Respiratory Medicine, Saiseikai Yokohamashi Nanbu Hospital, Yokohama, Japan. <sup>132</sup>Department of Clinical Laboratory, Saiseikai Yokohamashi Nanbu Hospital, Yokohama, Japan. <sup>133</sup>Internal Medicine, Internal Medicine Center, Showa University Koto Toyosu Hospital, Tokyo, Japan. <sup>134</sup>Department of Respiratory Medicine, Japan Organization of Occupational Health and Safety, Kanto Rosai Hospital, Kawasaki, Japan. <sup>135</sup>Department of General Internal Medicine, Japan Organization of Occupational Health and Safety, Kanto Rosai Hospital, Kawasaki, Japan. <sup>136</sup>Ishikawa Prefectural Central Hospital, Kanazawa, Japan. <sup>137</sup>Kanagawa Cardiovascular and Respiratory Center, Yokohama, Japan. <sup>138</sup>Department of Respiratory Medicine, National Hospital Organization Tokyo Medical Center, Tokyo, Japan. <sup>139</sup>Department of Allergy, National Hospital Organization Tokyo Medical Center, Tokyo, Japan. <sup>140</sup>Department of General Internal Medicine and Infectious Diseases, National Hospital Organization Tokyo Medical Center, Tokyo, Japan. <sup>141</sup>Department of Respiratory Medicine, Toyohashi Municipal Hospital, Toyohashi, Japan. <sup>142</sup>Keiyu Hospital, Yokohama, Japan. <sup>143</sup>KKR Sapporo Medical Center, Department of respiratory medicine, Sapporo, Japan. <sup>144</sup>Division of General Internal Medicine, Department of Internal Medicine, St. Marianna University School of Medicine, Kawasaki, Japan. <sup>145</sup>Department of Emergency and Critical Care Medicine, St. Marianna University School of Medicine, Kawasaki, Japan. <sup>146</sup>Japanese Red Cross Medical Center, Tokyo, Japan. <sup>147</sup>Matsumoto City Hospital, Matsumoto, Japan. <sup>148</sup>Department of Emergency and Critical Care Medicine, Faculty of Medicine, Fukuoka University, Fukuoka, Japan. <sup>149</sup>Department of Infection Control, Fukuoka University Hospital, Fukuoka, Japan. <sup>150</sup>Department of Rheumatology, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan. <sup>151</sup>Department of Respiratory Medicine, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan. <sup>152</sup>Department of Emergency and Critical Care Medicine, National Hospital Organization Hokkaido Medical Center, Sapporo, Japan. <sup>153</sup>National Hospital Organization Kanazawa Medical Center, Kanazawa, Japan. <sup>154</sup>Nihon University School of Medicine, Department of Internal Medicine, Division of Respiratory Medicine, Tokyo, Japan. <sup>155</sup>Musashino Red Cross Hospital, Musashino, Japan. <sup>156</sup>Division of Respiratory Medicine, Social Welfare Organization Saiseikai Imperial Gift Foundation, Inc., Saiseikai Kumamoto Hospital, Kumamoto, Japan. <sup>157</sup>Fukuoka Tokushukai Hospital, Department of Internal Medicine, Kasuga, Japan. <sup>158</sup>Fukuoka Tokushukai Hospital, Department Respiratory Medicine, Kasuga, Japan. <sup>159</sup>Department of Endocrinology, Hematology and Gerontology, Chiba University Graduate School of Medicine, Chiba, Japan. <sup>160</sup>Department of Emergency and Critical Care Medicine, Chiba University Graduate School of Medicine, Chiba, Japan. <sup>161</sup>National Hospital Organization Kumamoto Medical Center, Kumamoto, Japan. <sup>162</sup>Division of Infectious Diseases and Respiratory Medicine, Department of Internal Medicine, National Defense Medical College, Tokorozawa, Japan. <sup>163</sup>Sapporo City General Hospital, Sapporo, Japan. <sup>164</sup>Department of Internal Medicine, Tokyo Saiseikai Central Hospital, Tokyo, Japan. <sup>165</sup>Department of Pulmonary Medicine, Tokyo Saiseikai Central Hospital, Tokyo, Japan. <sup>166</sup>Department of Respiratory Medicine, Fujisawa City Hospital, Fujisawa, Japan. <sup>167</sup>Uji-Tokushukai Medical Center, Uji, Japan. <sup>168</sup>Department of Infectious Disease and Clinical Research Institute, National Hospital Organization Kyushu Medical Center, Fukuoka, Japan. <sup>169</sup>Department of Respirology, National Hospital Organization Kyushu Medical Center, Fukuoka, Japan. <sup>170</sup>Division of Respirology, Rheumatology, and Neurology, Department of Internal Medicine, Kurume University School of Medicine, Kurume, Japan. <sup>171</sup>Department of Infectious Disease, National Hospital Organization Kyushu Medical Center, Fukuoka, Japan. <sup>172</sup>Ome Municipal General Hospital, Ome, Japan. <sup>173</sup>Research Institute for Diseases of the Chest, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. <sup>174</sup>Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan. <sup>175</sup>Daini Osaka Police Hospital, Osaka, Japan. <sup>176</sup>Department of Emergency and Critical Care Medicine, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. <sup>177</sup>Department of Hematology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. <sup>178</sup>Department of Nephrology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. <sup>179</sup>Department of Cardiovascular Surgery, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. <sup>180</sup>Division of Pulmonary Medicine, Department of Medicine, Tokai University School of Medicine, Isehara, Japan. <sup>181</sup>Department of Anesthesiology and Intensive Care Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan. <sup>182</sup>Department of Infection Control and Laboratory Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan. <sup>183</sup>Department of Respiratory Internal Medicine, St. Marianna University School of Medicine, Yokohama-City Seibu Hospital, Yokohama, Japan. <sup>184</sup>Kinshukai Hanwa The Second Hospital, Osaka, Japan. <sup>185</sup>Gifu University School of Medicine Graduate School of Medicine, Emergency and Disaster Medicine, Gifu, Japan. <sup>186</sup>JA Toride Medical Hospital, Toride, Japan. <sup>187</sup>Okayama Rosai Hospital, Okayama, Japan. <sup>188</sup>Himeji St. Mary's Hospital, Himeji, Japan. <sup>189</sup>Emergency & Critical Care, Niigata University, Niigata, Japan. <sup>190</sup>Emergency & Critical Care Center, National Hospital Organization Kyoto Medical Center, Kyoto, Japan. <sup>191</sup>National Hospital Organization Tokyo Hospital Hospital, Kiyose, Japan. <sup>192</sup>Fujioka General Hospital, Fujioka, Japan. <sup>193</sup>Department of General Medicine, School of Medicine, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan. <sup>194</sup>Department of Pharmacology, School of Pharmacy, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan. <sup>195</sup>Department of Respiratory Medicine, International University of Health and Welfare Shioya

Hospital, Ohtawara, Japan.<sup>196</sup>Department of Clinical Laboratory, International University of Health and Welfare Shioya Hospital, Ohtawara, Japan.  
<sup>197</sup>Department of Cardiology, Pulmonology, and Nephrology, Yamagata University Faculty of Medicine, Yamagata, Japan. <sup>198</sup>Division of Respiratory  
Medicine and Allergology, Department of Medicine, School of Medicine, Showa University, Tokyo, Japan. <sup>199</sup>Kansai Electric Power Hospital, Osaka, Japan.  
<sup>200</sup>Division of Infectious Diseases, Kumamoto City Hospital, Kumamoto, Japan. <sup>201</sup>Department of Respiratory Medicine, Kumamoto City Hospital,  
Kumamoto, Japan. <sup>202</sup>Department of Emergency and Critical Care Medicine, Tokyo Metropolitan Police Hospital, Tokyo, Japan. <sup>203</sup>Department of  
Respiratory Medicine, Gunma University Graduate School of Medicine, Maebashi, Japan. <sup>204</sup>National Hospital Organization Saitama Hospital, Wako,  
Japan. <sup>205</sup>Tokyo Medical University Ibaraki Medical Center, Inashiki, Japan. <sup>206</sup>Department of Internal Medicine, Kiryu Kosei General Hospital, Kiryu, Japan.  
<sup>207</sup>Department of Pulmonary Medicine and Oncology, Graduate School of Medicine, Nippon Medical School, Tokyo, Japan. <sup>208</sup>Division of Respiratory  
Medicine, Tsukuba Kinen General Hospital, Tsukuba, Japan. <sup>209</sup>Division of Respiratory Medicine, Department of Internal Medicine, Toho University Ohashi  
Medical Center, Tokyo, Japan. <sup>210</sup>Division of Anesthesiology, Department of Surgery Related, Kobe University Graduate School of Medicine, Kobe, Japan.  
<sup>211</sup>Genome Medical Science Project (Toyama), National Center for Global Health and Medicine, Tokyo, Japan. <sup>212</sup>Department of Biomolecular Engineering,  
Graduate School of Tokyo Institute of Technology, Tokyo, Japan. <sup>213</sup>Laboratory of Veterinary Infectious Disease, School of Veterinary Medicine, Kitasato  
University, Aomori, Japan. <sup>214</sup>Laboratory of Viral Infection, Department of Infection Control and Immunology, Ōmura Satoshi Memorial Institute &  
Graduate School of Infection Control Sciences, Kitasato University, Tokyo, Japan. <sup>215</sup>Department of Insured Medical Care Management, Tokyo Medical and  
Dental University Hospital of Medicine, Tokyo, Japan. <sup>216</sup>Department of Organoid Medicine, Keio University School of Medicine, Tokyo, Japan.  
<sup>217</sup>Department of Surgery, Keio University School of Medicine, Tokyo, Japan.

## Methods

### Ethics

We have complied with all relevant ethical regulations. This study was approved by the ethical committees of Osaka University Graduate School of Medicine and affiliated institutes. All the participants provided written informed consent, as approved by the ethical committees of the affiliated institutes (Osaka University, Kyoto University and Keio University). No compensation was provided to the participants.

### Study populations

We enrolled 2,401 Japanese individuals comprising 1,062 PV, 378 RA, 251 SLE, 216 PAP, 36 MS, 94 COVID-19 patients and 364 healthy controls at Osaka University and related institutes. We also used CRAM files from WGS of 2,126 healthy controls generated by the Nagahama Study<sup>52</sup> and related studies at Kyoto University and BAM files from WGS of 1,070 patients with COVID-19 generated by the Japan COVID-19 Task Force<sup>53</sup>. For all patients, blood was drawn when the disease was active. Especially for patients with COVID-19, blood samples were drawn immediately after diagnosis. Also, we used the genomic DNA of 295 patients with PV and 429 healthy controls obtained from EBV-transformed B-lymphoblast cell lines of unrelated Japanese individuals established by the Japan Biological Informatics Consortium (JBIC)<sup>54</sup>.

We conducted principal component analysis (PCA) against HapMap3 data<sup>55</sup> using SNP data of the same individuals to confirm the East Asian genetic background.

### WGS data preprocessing

Genomic DNA was extracted from the whole blood of individuals at Osaka University and related institutes, with all samples processed at Osaka University. The DNA samples from Osaka University and those from the JBIC cell lines were sequenced at MacroGen Japan Corporation. Sequencing libraries were prepared using reagents from the TruSeq DNA PCR-free Library Prep kit (Illumina) following the manufacturer's protocols. Also, 2× 150 bp paired-end WGS was conducted on HiSeqX or NovaSeq (Illumina). The BWA-MEM (v0.7.13 or v0.7.8)<sup>56</sup> was used to align the sequenced reads to the human reference build hs37d5 (ref. 57). The reads in the raw alignments were flagged for duplication using Picard MarkDuplicates (v2.10.10 or v2.17.2). The number of average aligned reads was 386 million per sample, corresponding to an average genomic coverage of 18×. WGS CRAM files from the Nagahama Study and related studies contained 370 million 2× 150 bp paired-end reads per sample, aligned to hs37d5 with BWA-MEM (v0.7.13), and followed by Picard MarkDuplicates (v2.10.10), with a mean coverage of 17×. WGS BAM files from the Japan COVID-19 Task Force contained 751 million 2× 150 bp paired-end reads per sample, aligned to hs38DH<sup>58</sup> with BWA-MEM (v0.7.15), and followed by GATK4 MarkDuplicates<sup>59</sup> (v4.1.0.0), with a mean coverage of 26×. In total, 6,321 BAM or CRAM files were prepared for downstream analyses.

### Detection of eHHV-6A/eHHV-6B from WGS

The 6,321 WGS datasets were analyzed to detect eHHV-6 using a previously developed bioinformatics pipeline for eHHV-6 detection and reconstruction<sup>11</sup>. Briefly, we extracted reads that did not map to the human reference genome and mapped these reads to the reference HHV-6A (National Center for Biotechnology Information (NCBI) accession [NC\\_001664.4](https://www.ncbi.nlm.nih.gov/nuccore/NC_001664.4)) and HHV-6B ([NC\\_000898.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_000898.1)), followed by flagging for duplication to remove PCR duplicates. Then, the mapping depth and coverage of each virus were calculated. When 5% of the HHV-6 genome was covered at more than 2× read depth, we considered the sample to reflect the carriage of eHHV-6. Details are described in a previous paper<sup>11</sup>. The workflow described here was compiled as a Python pipeline and is available from the following GitHub repository: [https://github.com/shohei-kojima/integrated\\_HHV6\\_recon](https://github.com/shohei-kojima/integrated_HHV6_recon)

(<https://doi.org/10.5281/zenodo.12528938>)<sup>60</sup>. The 'main.py' in the repository wraps all the workflow described here.

### Detection of 'anellovirus reads' from WGS

The 5,597 WGS datasets were analyzed to detect reads derived from anelloviruses, excluding 724 participants from JBIC cell lines because cell lines are unlikely to reflect anellovirus infection status in the participants. Due to the extensive sequence divergence among the anellovirus genomes<sup>41,42</sup>, we modified our previously reported pipeline to accurately detect and quantify anelloviruses. We downloaded human anellovirus FASTA files from NCBI Virus on 21 April 2022. At this time, 27,780 anellovirus sequences were deposited in NCBI. To find human anelloviruses with complete genomes, we referred to the metadata of the deposited sequences and extracted 3,333 complete human anellovirus sequences that were annotated as follows: 'host', '*Homo sapiens*' and 'Nuc\_completeness', 'complete'. We found that some anellovirus genomes contain simple repeat sequences and regions with high similarity to human transposons. To avoid potential mismapping of human sequences to those regions, we masked simple repeats and human transposons by RepeatMasker<sup>61</sup> (v4.1.1) with '-species 9606 -s -no\_is' options. As a result, 1.09% of the anellovirus genomes were masked. The majority of the masked regions (103,757 bp) were simple repeats, but 69 and 1,724 bp were masked as long interspersed nuclear element-1 and long terminal repeat transposons, respectively, potentially reflecting mis-assemblies of viral genomes from metagenome sequencing data. We used the masked 3,333 genomes as anellovirus reference genomes. The human genome-unmapped reads were aligned to these 3,333 human anellovirus genomes using hisat2 (ref. 62) with '-mp 2,1 -no-spliced-alignment -all --secondary' options, followed by flagging for duplication using Picard MarkDuplicates. Then, the read pairs mapped to at least one anellovirus sequence were counted as 'Anellovirus reads'. Read pairs that were flagged as duplicates were excluded from 'Anellovirus reads'. We considered a read pair to be an 'Anellovirus read' when at least one of the read pairs was mapping to at least one anellovirus genome. The workflow described here was compiled as a Python pipeline and is available from the following GitHub repository: [https://github.com/shohei-kojima/human\\_anellovirus\\_detection](https://github.com/shohei-kojima/human_anellovirus_detection) (<https://doi.org/10.5281/zenodo.12528938>)<sup>60</sup>. The 'main.py' in the repository wraps the workflow described here.

To evaluate the mismapping of human sequences, we mapped 'Anellovirus reads' in 2,715 participants (cohorts excluding COVID-19) to CHM13. We found no mismapping of human reads to anellovirus genomes in a majority of the samples ( $n = 2,627$ ). In 88 samples (3.2% of participants), we observed mismapping of human reads to anelloviruses, but the number of mismappings was at most one read. This indicates that mismapping of human sequences (potentially nonreference sequences) is extremely rare, but a small amount is still present.

### Calculation of anellovirus load

We calculated anellovirus load (anellovirus copies per cell) to normalize anellovirus reads by WGS coverage as below:

$$\text{Anellovirus load} = \left( \frac{\text{anellovirus reads}}{\text{anellovirus genome length}} \right) \div \left( \frac{\text{human reads}}{\text{diploid human genome length}} \right)$$

'Anellovirus genome length' is the mean sequence length of 2,907.3 bp, calculated from 3,333 anellovirus FASTA files. 'Human reads' is the number of properly paired read pairs obtained using the SAMtools flagstat<sup>63</sup>. 'Diploid human genome length' is 5,618,659,254 bp in hs37d5 and 5,777,320,084 bp in hs38DH, calculated as the sum of 22 autosomal pairs and sex chromosome lengths of the human reference genome excluding Ns and gaps. The sex of the participants was not considered in this calculation.

### Thresholds for anellovirus infection states: positive, infection and viremia

Considering the possibilities of human-derived reads mappable on anellovirus sequences and index hopping, we did not simply define anellovirus-positive in the situation when a single (or small number of) ‘anellovirus read’ was detected. Therefore, we robustly defined ‘anellovirus reads’ >5 (anellovirus load > 0.08) as anellovirus-positive (Extended Data Fig. 6). To be sure, we tentatively loosened the threshold a little and evaluated ‘anellovirus reads’ >3 (anellovirus load > 0.05) as anellovirus-positive, but the conclusion remained the same (Supplementary Table 7). When anellovirus load = 1, there is one dsDNA genome per cell on average; therefore, participants with anellovirus load >1 were defined as anellovirus-infected. The maximum value of anellovirus load in the 2,490 healthy controls was 7.4. Thus, we defined participants with anellovirus load >8 as those demonstrating an atypical state of anellovirus viremia.

### Statistical analysis

The logistic regression model was used to calculate the adjusted case-control OR with 95% CI for the risk of each disease with eHHV-6 or each anellovirus infection state. We also used a two-sided Fisher’s exact test to evaluate the hypothesis of no association between the prevalence of eHHV-6A/eHHV-6B and the proportions of each anellovirus infection state in each disease and healthy controls. We applied the Bonferroni correction to adjust the *P* value for multiple test corrections. Then, the multivariable logistic regression model was performed with age and sex as independent variables. Anellovirus viremia was undetected in patients with PV, PAP or MS; therefore, no regression model was used. Wald test *P* value < 0.05 and Fisher’s test mid-*P* value < 0.05 were considered significant. All analyses were conducted using R (v4.0.5) with the R package MASS (v7.3.58)<sup>64</sup> and exact2x2 (v1.6.6)<sup>65</sup>. In each disease with a high OR, the linear or logistic regression model was used to estimate the relationship between the clinical metrics in each disease and eHHV-6 or each anellovirus infection state, with age and sex as independent variables. When the clinical metrics were binary, we also used Fisher’s test. Under regression analysis, listwise deletion of cases with missing values was applied.

### Replication in All of Us

To find potential eHHV-6 carriers from biobank-scale WGS datasets, we developed a fast-screening method. This screening uses unmapped reads and detects HHV-6 reads by aligning with the reference genomes of HHV-6. It first extracts unmapped reads stored at the end of a BAM or CRAM file using the ‘samtools view’ command and converts reads to the fastq format. The number of unmapped reads varies between individuals and is influenced by mapping algorithms. To minimize the computational burden, our screening approach uses no more than 1 million unmapped reads by default. To avoid mismapping of telomere reads to the telomere-like repeats in the reference genomes of HHV-6, it filters out reads containing at least one unit of the telomere repeat (TAACCC). After filtering potential telomere reads, it aligns reads with the HHV-6A and HHV-6B reference genomes, NCBI accessions [NC\\_001664.4](https://www.ncbi.nlm.nih.gov/nuccore/NC_001664.4) and [NC\\_000898.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_000898.1), respectively, using hisat2 with the relaxed mismatch penalties ‘--mp 2,1 --no-spliced-alignment’. To reduce the computational burden, it treats paired-end reads as single-end reads when mapping to HHV-6. Then, it counts reads successfully mapped to HHV-6. To avoid counting non-HHV-6 reads partially aligned with HHV-6, it counts reads without clipping and insertion-deletion (InDels). Finally, it outputs the numbers of reads screened and those mapping to HHV-6. This method is available as a script ‘quick\_check.py’ deposited in GitHub—[https://github.com/shohei-kojima/integrated\\_HHV6\\_recon](https://github.com/shohei-kojima/integrated_HHV6_recon) (<https://doi.org/10.5281/zenodo.12528938>)<sup>60</sup>. Because WGS reads were mapped to the human reference genome by DRAGEN, most of the CRAM files contain more than 1 million unmapped reads (usually around 15 million unmapped reads). To reduce the

computational burden, we used 1 million reads stored at the end of CRAM files for screening. In total, we screened 245,472 datasets in the All of Us Controlled Tier Dataset v7. For all datasets with six or more HHV-6 reads using the screening method, we applied the HHV-6 reconstruction method, ‘main.py’ in the integrated\_HHV6\_recon package to detect individuals carrying eHHV-6.

To define participants with SLE, we used diagnosis by the International Classification of Diseases, Ninth Revision (ICD9) (710.0) and ICD10 (M32) codes. To improve phenotype credibility, participants having two or more diagnoses on unique dates that were at least 1 year apart were defined as cases. As for PAP, another autoimmune disease with significant eHHV-6B risk in the Japanese population, we did not conduct a replication study given its ultra-rare disease prevalence (<10 cases per million population worldwide)<sup>66</sup>. We used all available participants for the association test, except for ones carrying ‘solo-DR’ and presumably two or more copies of full-length eHHV-6. We used the logistic regression model below. To avoid high type I errors, we performed a score test using saddle point approximation implemented in the R package SPAtest (<https://cran.r-project.org/web/packages/SPAtest/index.html>; v3.1.2).

$$\text{SLE} \sim \text{eHHV-6} + \text{age} + \text{age}^2 + \text{sex} + \text{top 10 genetic PCs.}$$

### GWAS of eHHV-6B positivity

From each disease, we randomly selected ten times as many eHHV-6B-negative participants as eHHV-6B-positive cases (22 eHHV-6B-positive versus 220 eHHV-6B-negative). The FASTQ reads were aligned to T2T-CHM13v2.0 (ref. 67) with BWA-MEM (v0.7.27), followed by GATK4 MarkDuplicates and Base Quality Score Recalibration (v4.2.6.1) according to the GATK Best Practice. Then, we performed per-sample SNP and InDel calling using GATK4 HaplotypeCaller and joint genotyping using GATK4 GenomicsDBImport and GenotypeGVCF. We conducted LD-based genotype refinement for low-confidence genotypes and missing sites in WGS data using BEAGLE (v5.4)<sup>68</sup> with default settings. Variant quality control (QC) was accomplished by keeping autosomal SNPs and InDels with minor allele frequency (MAF) ≥ 0.05, while removing multi-allelic sites, spanning deletions and variants deviating from the Hardy–Weinberg equilibrium ( $P < 1 \times 10^{-10}$  in cases and  $P < 1 \times 10^{-6}$  in controls). Individuals were excluded if they showed conflicting sex assignments between genetically inferred sex by variants and WGS coverage, deviating heterozygosity rate ( $\pm 3$  s.d.) or cryptic relatedness ( $\pi\text{-hat} > 0.2$ ). Finally, 6,464,509 variants and 238 participants (22 eHHV-6B-positive versus 216 eHHV-6B-negative) passed QC.

For GWAS, logistic regression analysis was conducted using PLINK (v2.0)<sup>69</sup> to test the differences in the dominant genotype between eHHV-6B and controls. We used the dominant genotype because eHHV-6B was designated as binary without distinguishing the number of alleles.

$$\text{eHHV-6B} \sim \text{variant} + \text{sex} + \text{PC1} + \text{PC2.}$$

Focusing on the distal end of chr22q, we performed logistic regression including rare variants (MAF < 0.05). Variant QC other than MAF filtering, sample QC and PCs followed the GWAS described above.

### PacBio HiFi long-read WGS

DNA samples from four eHHV-6B-positive patients with SLE were sheared targeting the size of 20 kb using Megaruptor 3 (Diagenode). HiFi sequencing libraries were prepared using SMARTbell Express Template Prep Kit 2.0 according to the manufacturer’s protocols. The libraries were further size-selected electrophoretically using either the SageELF or Blue Pippin Systems (Sage Science). The libraries were sequenced on the Sequel II (Pacific Bioscience) system using the Sequel II Binding Kit 2.0 and Sequel II Sequencing Kit 2.0. Based on

the sequenced subreads, PacBio HiFi reads were generated using the circular consensus sequence tool in SMART Link v9.0.0 (Pacific Bioscience). The HiFi reads were aligned to T2T-CHM13v2.0 using pbmm2 (v1.13.1; <https://github.com/PacificBiosciences/pbmm2>). DeepVariant (v1.6.0)<sup>70</sup> was used for SNP and InDel calling, and WhatsHap (v2.2)<sup>71</sup> was used for haplotype phasing. Phased genotypes were visually inspected using Integrative Genomics Viewer (v2.8.9)<sup>72</sup>.

### HHV-6 peptide phage library design

Open reading frame (ORF) annotations for HHV-6A strain U1102 (NC\_001664.4 and KC465951.1) and HHV-6B strain Z29 (NC\_000898.1, AF157706.1 and MW536483.1) were downloaded from NCBI. In total, 88 nonredundant ORFs for HHV-6A and 108 for HHV-6B were annotated. These ORFs were translated into amino acid sequences and then split into 64 amino acid peptide fragments with 50 amino acid overlaps. Fragments shorter than 64 amino acids at the C-terminus of a protein were extended with poly-alanine to reach the 64 amino acid length. This process yielded a total of 5,745 peptide fragments. For a positive control for the PhIP-seq library, we used a total of 248 human microbiota peptide antigens commonly detected in healthy individuals, as obtained from a previous study<sup>73</sup>. Using ICOR<sup>74</sup>, the peptide sequences were reverse-translated into DNA sequences optimized for *Escherichia coli* codon usage. HindIII and XhoI sites within the codon-optimized sequences were replaced with codons frequently used in *E. coli* (frequency >5% for each amino acid) in such a way that the protein sequence would be preserved.

### DNA oligo library synthesis, cloning and phage library synthesis

DNA oligo pools encoding the peptide sequences of HHV-6 were synthesized by Twist Bioscience. Sequences of the DNA oligos are listed in Supplementary Data 1. DNA oligos were PCR-amplified, digested with HindIII and XhoI and subsequently cloned into the HindIII/XhoI site of the T7 Select 10-3b phage vector. A ligation reaction was performed using T4 DNA Ligase. The resultant phage genomes were then packaged into T7 bacteriophage using the T7 Select Packaging Kit and amplified for use in PhIP-seq according to the manufacturer's protocol.

### PhIP-seq

We adapted previously established PhIP-seq protocols, with modifications<sup>28,29</sup>. IgG concentrations in serum or plasma samples were quantified using the Human IgG ELISA Kit. To minimize nonspecific binding, 1.5 ml tubes and Dynabeads protein A/G were blocked overnight at 4 °C on a rotator with 3% BSA in TBS-T. Sera or plasma containing 5 µg IgG and the HHV-6/human microbiota phage library diluted to 2 × 10<sup>5</sup>-fold representation (2 × 10<sup>10</sup> plaque-forming units for a library representing 10<sup>5</sup> peptide antigen repertoires) were added to preblocked tubes. Three technical replicates were performed for each sample, alongside 12 technical replicates for a negative control lacking sera or plasma. Phage-antibody mixtures were incubated overnight at 4 °C on a rotator, and 20 µl of Dynabeads protein A/G mix was added to each reaction. Following immunoprecipitation at 4 °C for 1 h on a rotator, beads were washed three times with 1,500 µl of PhIP-seq wash buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl and 0.1% Tween 20) and resuspended in lysogeny broth with antibiotics. *E. coli* culture (BLT5403 strain) was then added to resuspended beads and incubated at 37 °C with shaking until lysis. Phage lysates made 0.5 M in NaCl were centrifuged, and the supernatants diluted to 2 × 10<sup>5</sup>-fold representation were used for the second round of immunoprecipitation. Immunoprecipitated beads were subsequently washed three times, resuspended and lysed at 95 °C for 10 min. Finally, the genomes of immunoprecipitated phages were amplified by PCR for Illumina sequencing. A total of 2,000 copies of spike-in DNA (enhanced green fluorescent protein sequence) for normalization were also included in each PCR reaction. Sequences of the spike-in DNA and primer sequences are listed in Supplementary

Data 2 and 3. Following the PCR, technical replicates from the same sample were pooled at equimolar concentration and size-selected using SPRIselect. DNA concentrations were determined by quantitative PCR using the KAPA Library Quantification Kit.

### PhIP-seq data analysis

Phage genomes were amplified and sequenced on the Illumina NextSeq platform for 2 × 150 bp paired-end sequencing, yielding at least 0.24 M spots per sample. Amplicons should contain unique oligo sequences flanking phage genome sequences. First, we counted reads containing the flanking phage sequences. We allowed one base mismatch between next-generation sequencing (NGS) reads and the flanking phage genome sequences. Across all samples, at least 84% of reads contained phage genome sequences flanked by unique oligo sequences, validating that most PCR amplicons are derived from phages. Next, we detected the unique oligo sequences from the reads carrying the flanking phage sequences. We allowed one base mismatch between NGS reads and target oligos. Reads 1 and 2 containing the same oligo sequence were assigned to that oligo. This process enabled the assignment of at least 74% of the total reads to the oligo library across all samples.

To check whether most of the phages displaying HHV-6 peptides are included in the phage library, we amplified and sequenced them by the Illumina NextSeq platform. The 4.4 M spots were obtained, and, on average, 713 reads were detected for each oligo. In total, 313 oligos were not abundantly detected; the read counts were less than 5% of the average (<36 reads). We excluded those oligos from the data analysis.

The immunoprecipitated phage counts were calculated based on the number of reads detecting the spike-in oligo. Across all samples, the number of reads detecting the spike-in oligo was enough (>14,000 reads) compared to the actual spike-in oligo copy number (that is, 2,000), ensuring the accuracy of estimation of immunoprecipitated phage copy numbers. The immunoprecipitated phage counts were then normalized by the abundance of each oligo in the phage library. For this normalization, we used the normalization factor given below:

$$\text{factor}_i = \frac{\sum_n C}{n \times C_i}$$

where  $n$  is the number of oligos (that is, 5,432),  $C_i$  is the read count of the  $i$ th oligo detected in the sequencing of the phage library and  $\text{factor}_i$  is the normalization factor for the  $i$ th oligo.

The PhIP-seq score was calculated for each 14 amino acid region by averaging over the technical replicates and over the log-transformed normalized immunoprecipitated phage counts encoding the 14 amino acid region. The phage oligo detection from the NGS reads, normalization and the calculation of the PhIP-seq score were done by a custom Python script.

The statistical association between the presence of eHHV-6 and the PhIP-seq score was assessed by linear regression. We fit a model below, and the  $P$  for the  $\beta$  estimate of eHHV-6 was calculated by the Wald test ( $H_0: \beta = 0$ ). The model was fitted by the Python statsmodels module.

$$\begin{aligned} \log\text{-transformed PhIP-seq score} &\sim \text{eHHV-6} + \text{age} \\ &+ \text{sex} + \text{assay date} + \text{hospital} + \text{immunosuppressant use} \end{aligned}$$

### scRNA-seq

Four eHHV-6B-positive patients with SLE and 11 eHHV-6B-negative patients with SLE who were close to them in age, sex and treatment history were selected. For both, blood was collected onto heparin tubes, and PBMCs were isolated using Leucosep (Greiner Bio-One) density gradient centrifugation according to the manufacturer's instructions.

Single-cell suspensions were processed through the 10X Genomics Chromium Controller following the protocol outlined in the Chromium Single Cell V(D)J Reagent Kits (v1.1 Chemistry) User Guide. Chromium Next GEM Single Cell 5' Library and Gel Bead Kit v1.1 (PN-1000167), Chromium Next GEM Chip G Single Cell Kit (PN-1000127) and Single Index Kit T Set A (PN-1000213) were applied during the process. Oil droplets of encapsulated single cells and barcoded beads (gel bead-in-emulsions (GEMs)) were subsequently reverse-transcribed in a Veriti Thermal Cycler (Thermo Fisher Scientific), resulting in cDNA tagged with a cell barcode and unique molecular index (UMI). Next, cDNA was then amplified to generate single-cell libraries according to the manufacturer's protocol. Quantification was made with an Agilent Bioanalyzer High-Sensitivity DNA assay (High-Sensitivity DNA Kit; Agilent, 5067-4626). Subsequently amplified cDNA was enzymatically fragmented, end-repaired and polyA-tagged. Cleanup/size selection was performed on amplified cDNA using SPRIselect magnetic beads (Beckman-Coulter; SPRIselect, B23317). Next, Illumina sequencing adapters were ligated to the size-selected fragments and cleaned up using SPRIselect magnetic beads. Finally, sample indices were selected and amplified, followed by a double-sided size selection using SPRIselect magnetic beads. Final library quality was assessed using an Agilent Bioanalyzer High-Sensitivity DNA assay. Samples were then sequenced on an Illumina NovaSeq 6000 in paired-end mode.

### scRNA-seq data processing

Raw sequencing data were first filtered and then mapped to a custom reference (GRCh38 + HHV-6A/HHV-6B) using Cell Ranger (v.7.0.1; 10x Genomics). Filtered expression matrices were used. Then, cell doublets were removed by scds (v1.10.0)<sup>75</sup> for each sample. QC was performed to remove low-quality cells with <200 genes expressed, >6,000 genes expressed, >60,000 UMIs, < 5% of reads from ribosomal genes or >10% of reads from hemoglobin genes. Thresholds for filtering out low-quality cells with a high proportion of mitochondrial reads were determined using miQC (v1.2.0)<sup>76</sup>.

The R package Seurat (v4.3.0) was used for data scaling, transformation, clustering, dimensionality reduction and visualization<sup>32</sup>. Data were scaled and transformed using the SCTransform() function. Then, the first 2,000 highly variable genes were selected for PCA on gene expression using the SelectIntegrationFeatures() function. The batch effect among different patients was removed by the Harmony (v1.0)<sup>77</sup>. Uniform Manifold Approximation and Projection (UMAP) dimension reduction was generated based on the first 30 harmony-adjusted principal components. The nearest neighbor graph was calculated using the FindNeighbors() function, followed by clustering using the FindClusters() function. Cell types were annotated using the Azimuth human PBMC reference<sup>32,78</sup>.

### DE testing and GO enrichment

We used Milo (v1.2.0)<sup>33</sup> to test for the DE genes in cells within defined neighborhoods between the following two conditions: eHHV-6B-positive with SLE (cases) and eHHV-6B-negative with SLE (controls). We constructed a  $k$ -nearest neighbor graph and defined the representative neighborhoods on the  $k$ -nearest neighbor graph with default parameters. Then, we grouped overlapping neighborhoods that shared at least five cells and ignored their differences in abundance, although there was almost no difference. The neighborhood groups in which the number of cells in one case accounted for more than 60% of the total number of cells in the four cases were excluded from further analysis to avoid sampling bias. We performed DE analysis within these aggregated groups using the testDiffExp() function. We included age and sex in the model as covariates. Statistically significant DE genes were defined with  $FDR < 0.01$  and  $|\text{fold change}| > 2$ . To find the function of DE genes, GO biological process analysis with a  $P$  value  $< 0.01$  and an  $FDR < 0.05$  was performed using the enrichGO() and compareCluster() implemented in clusterProfiler (v4.7.1.3)<sup>79</sup>.

### Antiviral ISGs scores

To evaluate IFN response to the virus, antiviral ISG scores were calculated using the AddModuleScore() function implemented in Seurat with default parameters. Thirty-three antiviral ISGs were selected based on previous reports<sup>80–82</sup> (Supplementary Table 8). The ISGs scores of cells in each neighborhood were used to compare case–control, and the differences between their respective mean of the ISG scores in cells from two conditions were calculated. If there were two or more case samples and two or more control samples with cells in the neighborhood, Welch's  $t$  test was performed. To control for multiple testing, we adapted the spatial FDR implemented in Milo and used 5% spatial FDR as a threshold for significance. The spatial FDR and the differences in the ISG scores between two conditions in each neighborhood were used for visualization. Two-sided paired  $t$  test was performed to compare the mean of the ISG scores between two conditions in the neighborhoods within each neighborhood group.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Individual-level eHHV-6 and anellovirus data of the cases are publicly deposited at the Japanese Genotype–phenotype Archive (JGA) with accession codes JGAS000741/JGAD000876 and JGAS000739/JGAD000874 under controlled access. These individual-level eHHV-6 and anellovirus data can also be accessed through an application at the National Bioscience Database Center (NBDC) with the accession codes hum0197 and hum0343. The GWAS summary statistics are publicly deposited at the NBDC with the accession code hum0197. The WGS data of the controls are available from the Human Genetic Variation Database (HGVD) (<https://www.hgvd.genome.med.kyoto-u.ac.jp/>). The All of Us data are available as a featured workspace to registered researchers of the All of Us Researcher Workbench (see details at <https://www.researchallofus.org/>). Short/long-read WGS of eHHV-6B-positive participants are publicly deposited at the European Genome–Phenome Archive (EGA) with the accession code EGAS00001007886 under controlled access. PhIP-seq data of eHHV-6B-positive and eHHV-6B-negative participants with SLE are publicly deposited at EGA with the accession code EGAS00001007955 under controlled access. scRNA-seq data of eHHV-6B-positive and eHHV-6B-negative participants with SLE are publicly deposited at the JGA with accession codes JGAS000741/JGAD000876 under controlled access. This scRNA-seq data can also be accessed through the application at the NBDC with the accession code hum0197 (<https://humandbs.dbcls.jp/en/hum0197-latest>). The 3,333 human anellovirus complete genomes were downloaded as FASTA files from the NCBI nucleotide database (<https://www.ncbi.nlm.nih.gov>) on 21 April 2022.

### Code availability

The codes to detect virome, integrated\_HHV6\_recon and human\_anellovirus\_detection, and to perform WGS data preprocessing, GWAS, PhIP-seq analysis and scRNA-seq analysis are deposited in Zenodo (<https://doi.org/10.5281/zenodo.12528938> (ref. 60) and <https://doi.org/10.5281/zenodo.13953637> (ref. 83)).

### References

- Yano, M., Matsuda, F., Sakuntabhai, A., & Hirota, S. (eds). *Socio-Life Science and the COVID-19 Outbreak. Economics, Law, and Institutions in Asia Pacific* (Springer, 2022).
- Tanaka, H. et al. Clinical characteristics of patients with coronavirus disease (COVID-19): preliminary baseline report of Japan COVID-19 task force, a nationwide consortium to investigate host genetics of COVID-19. *Int. J. Infect. Dis.* **113**, 74–81 (2021).

54. Hirata, J. et al. Variants at HLA-A, HLA-C, and HLA-DQB1 confer risk of psoriasis vulgaris in Japanese. *J. Invest. Dermatol.* **138**, 542–548 (2018).
55. Altshuler, D. M. et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
56. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://doi.org/10.48550/arXiv.1303.3997> (2013).
57. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
58. Church, D. M. et al. Extending reference assembly models. *Genome Biol.* **16**, 13 (2015).
59. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
60. Kojima, S. Software of ‘Blood DNA virome associates with autoimmune diseases and COVID-19’. *Zenodo* <https://doi.org/10.5281/zenodo.12528938> (2024).
61. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. [www.repeatmasker.org](http://www.repeatmasker.org) (2013).
62. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
63. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
64. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S. Fourth Edition* Vol. 53 (Springer, 2002).
65. Fay, M. P. Confidence intervals that match Fisher’s exact or Blaker’s exact tests. *Biostatistics* **11**, 373–374 (2010).
66. Inoue, Y. et al. Characteristics of a large cohort of patients with autoimmune pulmonary alveolar proteinosis in Japan. *Am. J. Respir. Crit. Care Med.* **177**, 752–762 (2008).
67. Rhie, A. et al. The complete sequence of a human Y chromosome. *Nature* **621**, 344–354 (2023).
68. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
69. Hill, A. et al. Stepwise distributed open innovation contests for software development: acceleration of genome-wide association analysis. *GigaScience* **6**, 1–10 (2017).
70. Poplin, R. et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987 (2018).
71. Patterson, M. et al. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. *J. Comput. Biol.* **22**, 498–509 (2015).
72. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
73. Vogl, T. et al. Population-wide diversity and stability of serum antibody epitope repertoires against human microbiota. *Nat. Med.* **27**, 1442–1450 (2021).
74. Jain, R., Jain, A., Mauro, E., LeShane, K. & Densmore, D. ICOR: improving codon optimization with recurrent neural networks. *BMC Bioinformatics* **24**, 132 (2023).
75. Bais, A. S. & Kostka, D. scds: computational annotation of doublets in single-cell RNA sequencing data. *Bioinformatics* **36**, 1150–1158 (2020).
76. Hippen, A. A. et al. miQC: an adaptive probabilistic framework for quality control of single-cell RNA-sequencing data. *PLoS Comput. Biol.* **17**, e1009290 (2021).
77. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
78. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
79. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
80. Schoggins, J. W. & Rice, C. M. Interferon-stimulated genes and their antiviral effector functions. *Curr. Opin. Virol.* **1**, 519–525 (2011).
81. Schoggins, J. W. Interferon-stimulated genes: what do they all do? *Annu. Rev. Virol.* **6**, 567–584 (2019).
82. Chemudupati, M. et al. From APOBEC to ZAP: Diverse mechanisms used by cellular restriction factors to inhibit virus infections. *Biochim. Biophys. Acta Mol. Cell Res.* **1866**, 382–394 (2019).
83. Sasa, N. & Kojima, S. Scripts for ‘Blood DNA virome associates with autoimmune diseases and COVID-19’. *Zenodo* <https://doi.org/10.5281/zenodo.13953637> (2024).

## Acknowledgements

We thank all the participants involved in this study. We gratefully acknowledge All of Us participants for their contributions, without whom this research would not have been possible. We also thank the NIH’s All of Us Research Program for making available the participant data examined in this study. This research was supported by the Japan Society for the Promotion of Science KAKENHI (22H00476 and 24K02697 to Y.O., 20H05682 and 21H02972 to N.F.P., 22K15385 to S. Kojima and 24K18476 to R. Koide), Japan Agency for Medical Research and Development (AMED) (JP22ek0410075, JP23km0405211, JP23km0405217, JP23ek0109594, JP23ek0410113, JP223fa627002, JP223fa627010, JP233fa627011, JP23zf0127008, JP24wm0625504 and JP24gm1810011 to Y.O. and JP24gm7010004 to S. Kojima), JST Moonshot R&D (JPMJMS2021 and JPMJMS2024 to Y.O.), Takeda Science Foundation, Bioinformatics Initiative of Osaka University Graduate School of Medicine, Institute for Open and Transdisciplinary Research Initiatives, Center for Infectious Disease Education and Research, Center for Advanced Modality and Drug Delivery System (DDS), Osaka University, Cluster for Pioneering Research under the Hakubi fellowship program, the discretionary budget of the Director of the RIKEN Center for Integrative Medical Sciences and the RIKEN Special Postdoctoral Researcher Program. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper.

## Author contributions

N.F.P. and Y.O. designed and supervised the study. N.S., S. Kojima and Y.O. conducted the data analysis. S. Kojima, N.S. and N.F.P. developed methodology. R. Koide performed PhIP-seq. N.S., S. Kojima, R. Koide, A.G.B., N.F.P. and Y.O. wrote the paper. All the other co-authors collected and managed the samples and data construction.

## Competing interests

The authors declare no competing interests.

## Additional information

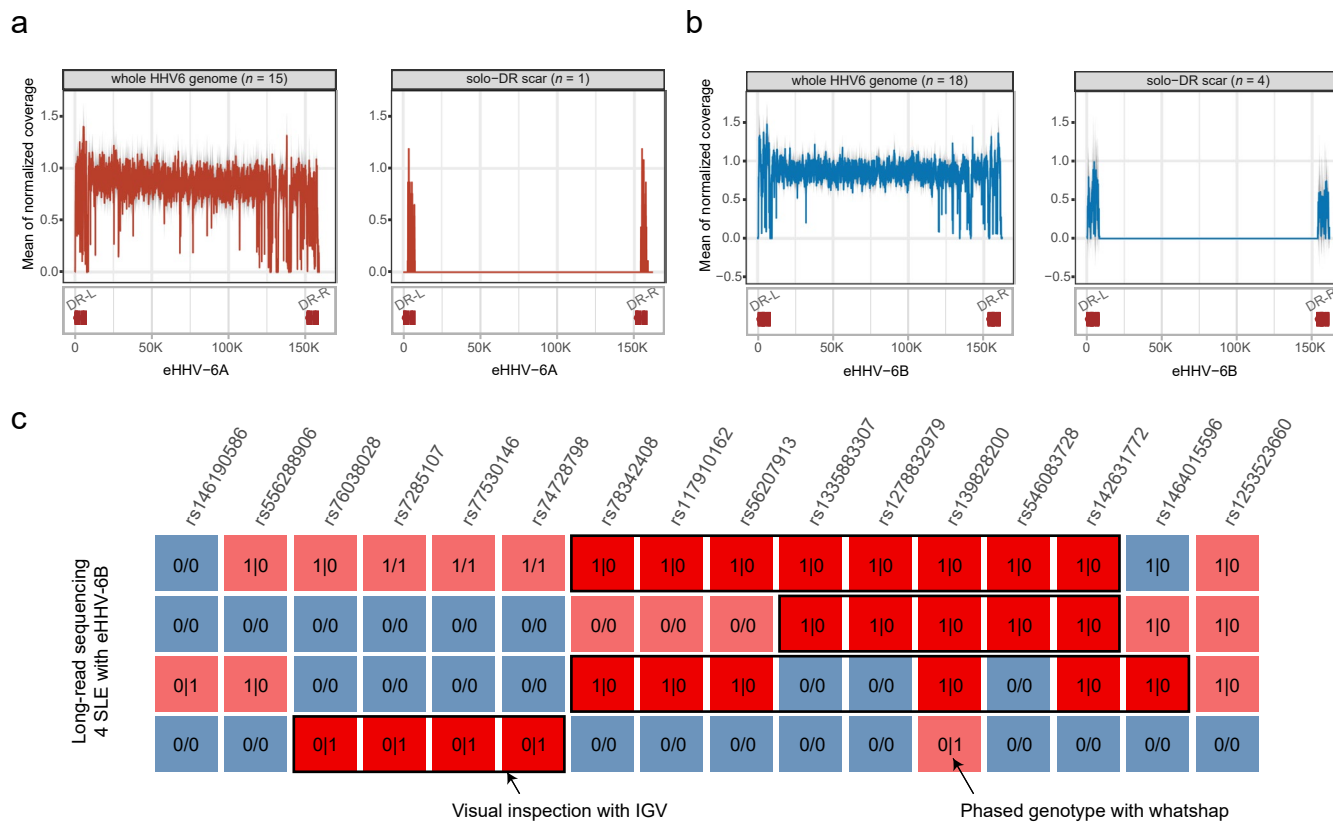
**Extended data** is available for this paper at <https://doi.org/10.1038/s41588-024-02022-z>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-024-02022-z>.

**Correspondence and requests for materials** should be addressed to Nicholas F. Parrish or Yukinori Okada.

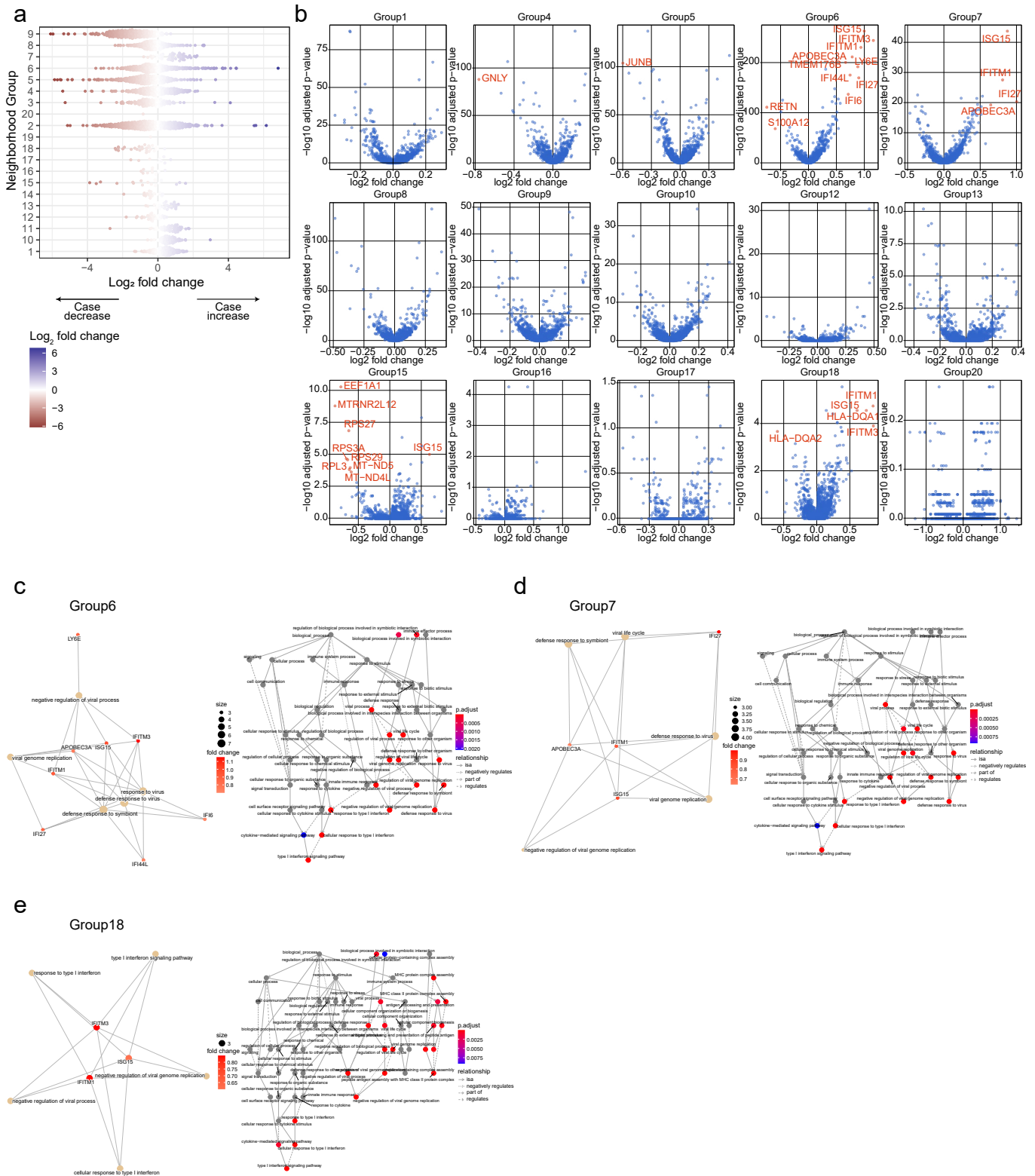
**Peer review information** *Nature Genetics* thanks Mike Nalls and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



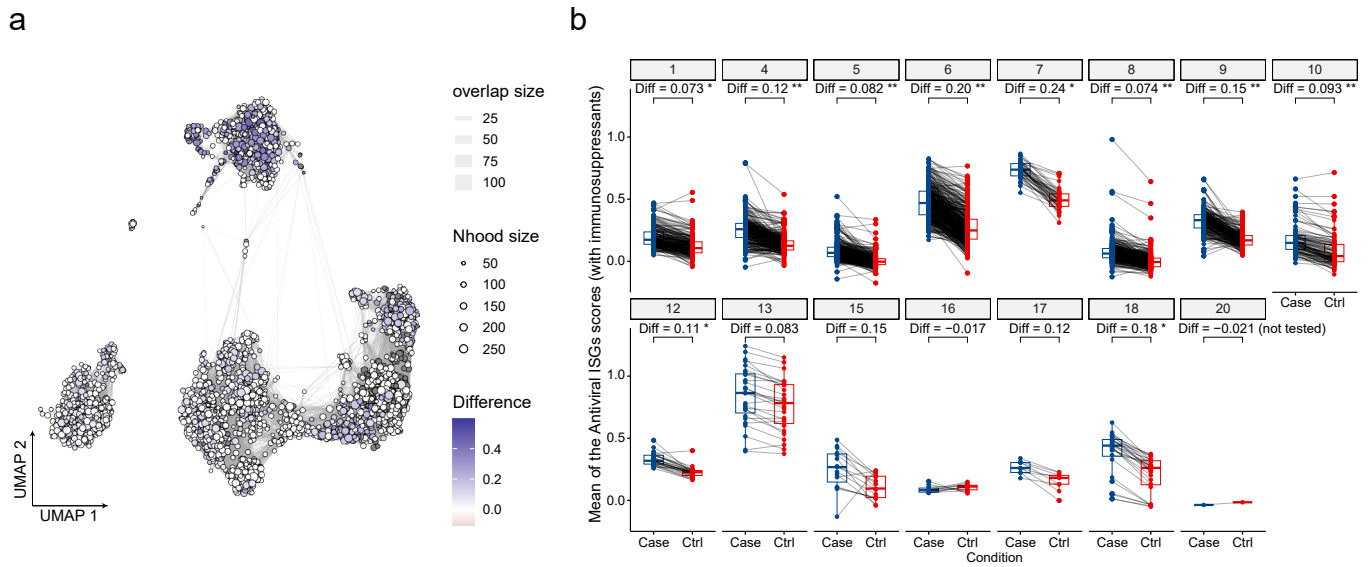
**Extended Data Fig. 1 | HHV-6 genome coverage and haplotypes on the distal end of chr22q in four eHHV-6B-positive patients with SLE. a, b.** The mean of HHV-6A/B genome coverages normalized by half of the human genome coverage. The gray shade area represents the 95% confidence interval. Typically, eHHV-6-positive subjects carry the whole HHV-6 genome, but some carry only one of the terminal direct repeats (DRs)—solo-DR—and not the unique (U) region of the HHV-6 genome<sup>9</sup>. All four eHHV-6B-positive patients with SLE carried the whole HHV-6B genome, and two out of three eHHV-6B-positive patients with PAP carried only solo-DR. The mean normalized coverage of full-length HHV-6

was approximately 1, which supports the HHV-6 genomes we found are heritable endogenous viruses. Since the aligner could not distinguish between DR-L and DR-R sequences, solo-DR-derived reads were mapped to both. The mean normalized coverage of solo-DR was approximately 0.5, which also supports the endogeneity of the HHV-6 genome. **c.** Phased genotypes of the 16 variants with uncorrected  $P < 1 \times 10^{-4}$  in GWAS, derived from long-read sequencing of 4 eHHV-6B-positive patients with SLE. Dark red and black boxes indicate haploid visually confirmed with IGV. Light red and blue indicate genotypes from WGS joint genotyping. IGV, Integrative Genomics Viewer.



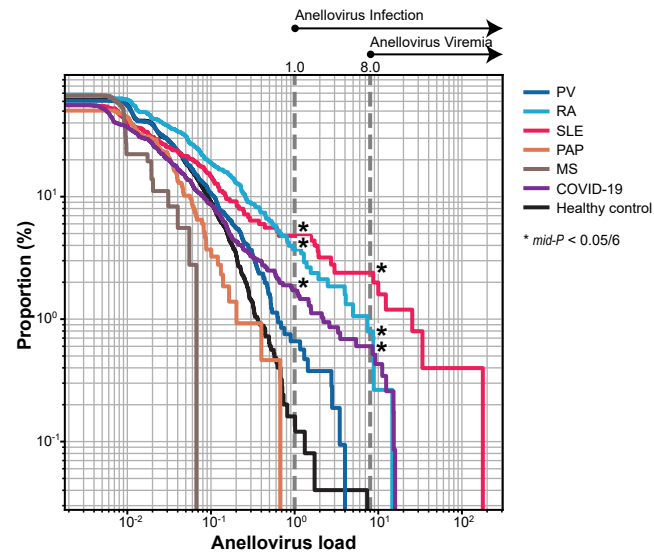
**Extended Data Fig. 2 | DA analysis, DE analysis and GO enrichment analysis of scRNA-seq data of PBMCs from eHHV-6B-positive and eHHV-6B-negative patients with SLE. a**, Beeswarm plot showing the distribution of adjusted  $\log_2(\text{fold change})$  in abundance between cases and controls in neighborhoods, with eHHV-6B-positive patients with SLE as case ( $n = 4$ ) and eHHV-6B-negative patients with SLE as control ( $n = 11$ ). **b**, The DE analysis between cases and controls in neighborhood groups as in **a**. DE genes ( $FDR < 0.01$  and  $|\text{fold change}| > 2$ ) are colored in red and labeled by gene symbols. **c–e**, Left: gene-

concept network (cneplot) in neighborhood groups 6, 7 and 18. It shows the linkage of the DE genes and the concepts (GO biological process). The color of the genes represents the fold change values, and the size of the GO terms represents the number of associated genes. Right: directed acyclic graph for GO enrichment analysis in neighborhood groups 6, 7 and 18. Circles represent the GO function, colored by their FDR. Arrows indicate the relationship between the upstream and downstream GO functions.



**Extended Data Fig. 3 | Comparison of the antiviral ISG scores in patients with SLE on immunosuppressants. a**, Graph representation of neighborhoods identified by Milo as in Fig. 5b. Node color indicates the difference between the mean of the antiviral ISG scores of eHHV-6B-positive patients (case;  $n = 3$ ) and eHHV-6B-negative patients (control;  $n = 6$ ) with SLE on immunosuppressants (spatial FDR < 0.05). Neighborhoods with non-significant

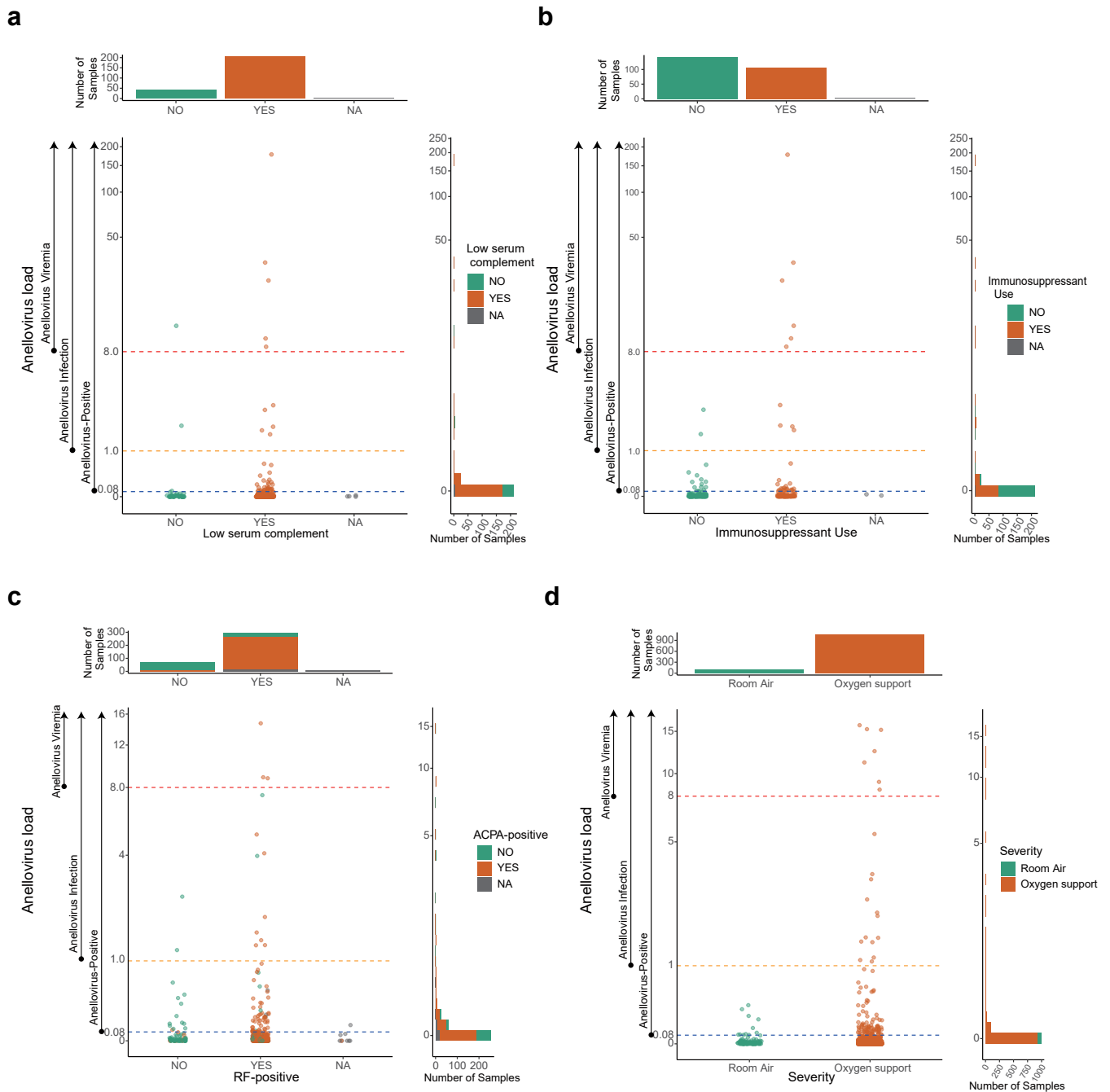
differences (spatial FDR  $\geq 0.05$ ) are colored white. Neighborhoods that contain only cells of either cases or controls are colored black. **b**, Comparison of differences within each neighborhood group shown in **a** using a two-sided paired  $t$ -test. Diff is the average difference between the case and control means. Boxes denote the IQR; the median is shown as horizontal bars; whiskers extend to 1.5 times the IQR. \*FDR <  $1.0 \times 10^{-10}$ , \*\*FDR <  $1.0 \times 10^{-50}$ .



**Extended Data Fig. 4 | Reverse cumulative distributions of anellovirus load.**

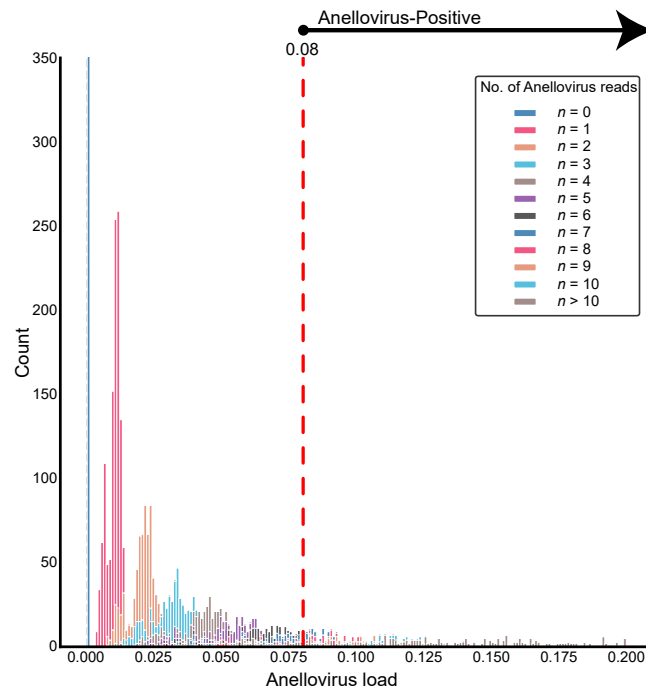
The reverse cumulative distribution plot of anellovirus load in subjects with five autoimmune diseases and COVID-19 and healthy controls. Y-axis indicates the proportions of subjects with anellovirus load > values (x-axis). The proportions of anellovirus infection with anellovirus load >1 were significantly higher in

patients with SLE, RA and COVID-19 than in healthy controls. The proportions of anellovirus viremia with anellovirus load >8.0 were significantly higher in patients with SLE, RA and COVID-19 than in healthy controls (dashed lines and asterisks).



**Extended Data Fig. 5 | Distribution of anellovirus load among various clinical metrics.** Two histograms show the distribution of each axis. **a**, Strip plot of anellovirus load for patients with SLE without and with low serum complement levels. Anellovirus-positive subjects associated significantly with low serum complement levels (OR 4.4, Wald test uncorrected  $P = 0.020$ ). NA, not available. **b**, Strip plot of anellovirus load for patients with SLE treated without and with immunosuppressants. Anellovirus-positive and anellovirus-infected subjects had a significantly positive correlation with immunosuppressant use (OR 3.3 and

7.3, Wald test uncorrected  $P = 0.0050$  and  $0.012$ , respectively). **c**, Strip plot of anellovirus load for patients with RA stratified by RF and ACPA. All anellovirus-viremic subjects were RF-positive and ACPA-positive. **d**, Strip plot of anellovirus load for patients with COVID-19 treated without and with oxygen therapy. All subjects with a high anellovirus load were treated with oxygen therapy. Horizontal dashed lines in **a–d** represent thresholds for anellovirus positivity (0.08), anellovirus infection (1.0) and anellovirus viremia (8.0).



**Extended Data Fig. 6 | The detection limit of the anellovirus load.** Histogram of anellovirus load below 0.2. The colors of the bars indicate the number of 'anellovirus reads' in each sample. We set the detection limit of the anellovirus

load to 0.08 (red dashed line). Subjects with 'anellovirus reads' >5 (anellovirus load >0.08) were defined as anellovirus-positive. Subjects with anellovirus load between 0 and 0.08 might be false-positive.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Individual-level eHHV-6 and anellovirus data of the cases are publicly deposited at the Japanese Genotype-phenotype Archive (JGA) with accession codes JGAS000741 (<https://ddbj.nig.ac.jp/resource/jga-study/JGAS000741/JGAD000876>) and JGAS000739 (<https://ddbj.nig.ac.jp/resource/jga-study/JGAS000739/JGAD000874>) under controlled access. These individual-level eHHV-6 and anellovirus data can also be accessed through application at the NBDC with the accession code hum0197 (<https://humandbs.dbcls.jp/en/hum0197-latest>) and hum0343 (<https://humandbs.dbcls.jp/en/hum0343-latest>). The GWAS summary statistics are publicly deposited at the NBDC with the accession code hum0197 (<https://humandbs.dbcls.jp/en/hum0197-latest>). The WGS data of the controls are available from HGVD (<https://www.hgvd.genome.med.kyoto-u.ac.jp/>). The All of Us data are available as a featured workspace to registered researchers of the All of Us Researcher Workbench (see details at <https://www.researchallofus.org/>). Short/long-read WGS of eHHV-6B-positive subjects are publicly deposited at the European Genome-Phenome Archive (EGA) with the accession code EGAS00001007886 (<https://ega-archive.org/studies/EGAS00001007886>) under controlled access. PhiP-seq data of eHHV-6B-positive and -negative subjects with SLE are publicly deposited at EGA with the accession code EGAS00001007955 (<https://ega-archive.org/studies/EGAS00001007955>) under controlled access. Single-cell RNA sequencing data of eHHV-6B-positive and -negative subjects with SLE are publicly deposited at the Japanese Genotype-phenotype Archive (JGA) with accession codes JGAS000741 (<https://ddbj.nig.ac.jp/resource/jga-study/JGAS000741/JGAD000876>) and JGAS000739 (<https://ddbj.nig.ac.jp/resource/jga-study/JGAS000739/JGAD000874>) under controlled access. This scRNA-seq data can also be accessed through application at the NBDC with the accession code hum0197 (<https://humandbs.dbcls.jp/en/hum0197-latest>). We manually extracted 3,333 complete human anellovirus FASTA files from NCBI Virus ([https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType\\_s=Nucleotide&VirusLineage\\_ss=Anelloviridae,%20taxid:687329&utm\\_source=data-hub](https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Anelloviridae,%20taxid:687329&utm_source=data-hub)).

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

### Reporting on sex and gender

Genetically determined sex was collected and included in linear or logistic regression analysis as a covariate.  
None of our findings applies to only one sex or gender.

### Population characteristics

Population characteristics of the human research participants that were included as covariates are summarized below. This information is also included in Supplementary Table 1.

Disease, Sample size, Sex N (F/M), Age (mean SD; for those were the data was available)

Association study for eHHV-6A/B:  
 PV, 1,357, 333 / 1,024, 54.25 (15.16)  
 RA, 378, 295 / 83, 66.07 (13.43)  
 SLE, 251, 221 / 30, 45.20 (16.09)  
 PAP, 216, 90 / 126, 56.49 (13.85)  
 MS, 36, 25 / 11, 40.36 (9.54)  
 COVID-19, 1,164, 323 / 841, 59.67 (16.37)  
 Healthy controls, 2,919, 1,789 / 1,130, 51.39 (16.63)

Association study for anellovirus infection:  
 PV, 1,062, 269 / 793, 54.90 (14.97)  
 RA, 378, 295 / 83, 66.07 (13.43)  
 SLE, 251, 221 / 30, 45.20 (16.09)  
 PAP, 216, 90 / 126, 56.49 (13.85)  
 MS, 36, 25 / 11, 40.36 (9.54)  
 COVID-19, 1,164, 323 / 841, 59.67 (16.37)  
 Healthy controls, 2,490, 1,595 / 895, 54.03 (16.25)

### Recruitment

We enrolled 2,401 Japanese individuals composed of 1,062 PV, 378 RA, 251 SLE, 216 PAP, 36 MS, 94 COVID-19 patients, and 364 healthy controls in Osaka University Hospital and related institutes. Part of the samples have been used in previously published papers (Sakaue, S. et al., Nat Commun, 2021; Tanaka, H. et al., Ann Rheum Dis, 2024; Tomofuji, Y. et al., Nat Microbiol, 2023; Namkoong, H. et al., Nature, 2022). We used the genomic DNA of 295 patients with PV and 429 healthy controls obtained from Epstein–Barr virus-transformed B-lymphoblast cell lines of unrelated Japanese individuals established by the Japan Biological Informatics Consortium (JBIC) as described in Hirata, J. et al., J Invest Dermatol, 2018. We also used CRAM files from WGS of 2,126 healthy controls generated by the Nagahama Study and related studies at Kyoto University (Setoh, K. et al., Economics, Law, and Institutions in Asia Pacific, 2022) and BAM files from WGS of 1,070 patients with COVID-19 generated by the Japan COVID-19 Task Force (Wang, Q. S. et al., Nat Commun, 2022).

### Ethics oversight

This study was approved by the ethical committee of Osaka University Graduate School of Medicine and affiliated institutes.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A total of 1,357 patients with PV, 378 patients with RA, 251 patients with SLE, 216 patients with PAP, 36 patients with MS, 1,164 patients with COVID-19, and 2,919 healthy controls were included in the analysis. The aim was to utilize the largest available dataset in order to maximise the power to detect associations, so no sample size calculation was performed.
Data exclusions	No data was excluded. We conducted principal component analysis (PCA) against HapMap3 data using the individuals' data to confirm the East Asian genetic background.
Replication	We conducted two large-scale replication studies of the association between SLE and eHHV-6B in European ancestry and multi-ancestry populations using All of Us WGS data, both of which were successful.
Randomization	Not applicable, this is an observational study.
Blinding	We did not apply blinding of the samples because this is a genotype-phenotype association study and no intervention was conducted in our study.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

### Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used      We did not use commercial antibodies. Therefore, catalog number, clone name, lot number, and validation data are not available.

Validation      Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.