scientific reports

OPEN

Mendelian randomization unveils genetic causal relationships between viral infections and systemic sclerosis

Ting Cheng[1,2,3], He-Yi Zhang[2,3], Tian-Run Wang[2,3], Ruo-Lan Li[2,3], Zhi-Nan Jing[2,3], Yue-Hong Huo[2,3,4✉] & Sheng-Xiao Zhang[1,2,3,5✉]

This study uses Mendelian randomization (MR) to investigate the potential causal relationships between viral infections, including Epstein-Barr virus (EBV), human immunodeficiency virus (HIV), SARS-CoV-2, cytomegalovirus (CMV), human herpes virus 6 (HHV-6), varicella-zoster virus (VZV), herpes simplex virus (HSV), influenza A virus, and hepatitis B virus, and the risk of systemic sclerosis (SSc). Summary-level data on viral exposures and SSc outcomes were obtained from public genome-wide association studies (GWAS) databases. Causality was assessed using the inverse-variance weighted (IVW), MR-Egger, and weighted median methods. Sensitivity analysis was conducted to enhance the reliability and robustness of our findings. Genetically predicted anti-EBV viral capsid antigen IgG levels (OR = 3.400, 95% CI = 1.093–10.571, $p = 0.035$) were causally associated with an elevated risk of SSc, while HIV (OR = 0.787, 95% CI = 0.629–0.985, $p = 0.037$) and SARS-CoV-2 (OR = 0.335, 95% CI = 0.116–0.964, $p = 0.043$) correlated with a reduced risk of SSc. Sensitivity analysis validated the robustness of these associations ($p > 0.05$). Further elucidation of the underlying mechanisms by which EBV increases the risk of SSc could potentially identify interventions for promoting SSc prevention.

Keywords Virus infection, Systemic sclerosis, Mendelian randomization, Causal relationship, Genome-wide association study

Systemic sclerosis (SSc) is a multifaceted autoimmune connective tissue disease characterized by vasculopathy, immune dysregulation, and progressive fibrosis that typically affects skin and internal organs, carrying a significant burden of disease-related morbidity including potentially life-threatening complications[1,2]. Despite extensive research efforts aimed at unraveling the underlying mechanisms of SSc, the exact etiology of this disease remains elusive. Recent years have witnessed a surge in interest concerning the potential role of viral infections in the development and progression of SSc[3].

Viruses can trigger autoimmunity through diverse mechanisms, including molecular mimicry, epitope spreading, bystander activation, and immortalization of infected B cells[4]. Patients with SSc have been observed to exhibit elevated rates of antibodies against various viruses, including cytomegalovirus (CMV), hepatitis B virus, and Epstein-Barr virus (EBV), in comparison to healthy controls[5]. A notable serological association between EBV, particularly in its reactive stage, and SSc was established in a cross-sectional study that hinged on the absence of viral DNA in the circulation[6]. Importantly, it is crucial to recognize that correlations observed in observational studies do not necessarily imply causation. Hence, a comprehensive exploration is necessary to validate the relationship between viral infections and SSc.

In this study, we employed Mendelian randomization (MR) to investigate the potential causal links between EBV, human immunodeficiency virus (HIV), SARS-CoV-2, CMV, human herpes virus 6 (HHV-6), varicella-zoster virus (VZV), herpes simplex virus (HSV), influenza A virus, hepatitis B virus and SSc. Elevated antibody levels typically signify an active stage of viral infection or reactivation of the virus[7–9]. This is why we used

[1]Department of Rheumatology, the Second Hospital of Shanxi Medical University, Taiyuan030001, China. [2]Shanxi Provincial Key Laboratory of Rheumatism Immune Microecology, Taiyuan, Shanxi Province, China. [3]Key Laboratory of Cellular Physiology, Ministry of Education, Shanxi Medical University, Taiyuan, Shanxi Province, China. [4]Department of Rheumatology, The Fifth People's Hospital of Datong, Datong 037000, Shanxi, China. [5]SXMU-Tsinghua Collaborative Innovation Center for Frontier Medicine, Shanxi Medical University, Taiyuan, Shanxi Province, China. ✉email: 63725438@qq.com; zhangshengxiao1@sxmu.edu.cn

elevated antibody levels as an indicator of virus infection. MR, as a statistical method relying on whole genome sequencing data, offers the advantage of being impervious to confounding factors and reverse causalities. It accomplishes this by utilizing genetic variants associated with exposures as proxies for the exposures themselves, facilitating the estimation of causal associations between these proxies and the ultimate outcome[10]. Furthermore, MR has gained widespread acceptance as a valuable approach for estimating the causal association between SSc and other factors[11–13].

Materials and methods
Study design and assumptions
The two-sample MR study was conducted following the guidelines for strengthening the reporting of observational studies in epidemiology using Mendelian Randomization (STROBE-MR)[14]. All genetic variants utilized as instrumental variables (IVs) to estimate causal effects must adhere to three fundamental hypotheses[1]: the correlation hypothesis, meaning that genetic variants are strongly associated with the exposure[2]; the independence hypothesis, indicating that genetic variants are not linked to any confounders affecting the exposure-outcome association; and[3] the exclusion restriction hypothesis, asserting that genetic variants influence the outcome solely through their impact on the exposure.

Exposure GWAS data sources
We obtained summary single nucleotide polymorphism (SNP)-phenotype association data for exposures from the Integrative Epidemiology Unit (IEU) open GWAS database (https://gwas.mrcieu.ac.uk/). The data were restricted to individuals of European descent. Specifically, the dataset for anti-EBV VCA IgG levels included 956 participants and 5,278,042 SNPs[15]. HIV disease data comprised 357 cases and 218,435 controls with 16,380,466 SNPs. Data for SARS-CoV-2 encompassed 1,683,768 samples (N = 38,984 cases, 1,644,784 controls) and 8,660,177 SNPs[16]. The dataset for anti-CMV IgG seropositivity featured 8,735 samples and 9,170,312 SNPs[17]. HHV-6 1E1A antibody levels data included 6,968 samples and 9,170,460 SNPs[17]. The dataset for varicella (chickenpox) (N = 710 cases, 211,856 controls) consisted of 16,380,433 SNPs, and the herpesviral (herpes simplex) infections data (N = 1,595 cases, 211,856 controls) comprised 16,380,457 SNPs. The public data for anti-influenza A virus IgG levels involved 777 samples and 5,278,042 SNPs[15]. Lastly, the data for anti-HBsAg IgG levels (N = 508) included 5,278,042 SNPs[15].

Outcome GWAS data sources
Summary-level data for SSc were derived from the FinnGen research project (GWAS ID: finn-b-SYSTSCLE_STRICT), consisting of 107 patients with SSc and 218,499 healthy controls, all of European ancestry. All cases were diagnosed by medical institutions and met at least one of the following criteria: (1) Meeting the 2013 ACR/EULAR classification criteria for systemic sclerosis; (2) Having a diagnosis record of SSc in a specialist outpatient clinic (e.g., rheumatology and immunology department) (ICD-10 code: M34); (3) Presenting with characteristic SSc antibodies (such as anti-topoisomerase I or anti-centromere antibodies) and typical clinical manifestations (such as Raynaud's phenomenon, skin fibrosis, or visceral involvement). (https://finngen.gitbook.io/documentation/) Key details of the data are presented in Table 1.

Selection of IVs
A series of quality control steps were performed to identify eligible IVs. Firstly, SNPs should be significantly associated with the exposures and satisfy genome-wide significance thresholds ($p < 1 \times 10^{-5}$). The F-statistic was calculated to assess the strength of the selected SNPs, and only those with an F-statistic greater than 10 were retained as effective IVs. Then, the influence of linkage disequilibrium (LD) was mitigated by setting a threshold ($r^2 < 0.001$, 1000 kb) to keep independence among IVs for each exposure. In addition, palindrome and incompatible SNPs were excluded by harmonizing the exposure and outcome data. Finally, we utilized the PhenoScannerV2 database (http://www.phenoscaner.medschl.cam.ac.uk/) to manually screen and remove SNPs related to confounding factors, such as smoking, alcohol consumption, and body mass index (BMI).

Trait	GWAS ID	Cases	Controls	Sample size	SNPs	Year	Population
Anti-Epstein-Barr virus viral capsid antigen IgG levels[a]	ebi-a-GCST006362	NA	NA	956	5,278,042	2018	European
Human immunodeficiency virus (HIV) disease[a]	finn-b-AB1_HIV	357	218,435	218,792	16,380,466	2021	European
SARS-CoV-2[a]	ebi-a-GCST011073	38,984	1,644,784	1,683,768	8,660,177	2020	European
Anti-cytomegalovirus IgG seropositivity[a]	ebi-a-GCST90006893	NA	NA	8,735	9,170,312	2020	European
Human herpes virus 6 1E1A antibody levels[a]	ebi-a-GCST90006904	NA	NA	6,968	9,170,460	2020	European
Varicella[a]	finn-b-AB1_VARICELLA	710	211,856	211,566	16,380,433	2021	European
Herpesviral infections[a]	finn-b-AB1_HERPES_SIMPLEX	1,595	211,856	213,451	16,380,457	2021	European
Anti-influenza A virus IgG levels[a]	ebi-a-GCST006350	NA	NA	777	5,278,042	2018	European
Anti-hepatitis B virus surface antigen IgG levels[a]	ebi-a-GCST006354	NA	NA	508	5,278,042	2018	European
Systemic sclerosis[b]	finn-b-SYSTSCLE_STRICT	107	218,499	218,606	16,380,462	2021	European

Table 1. The details of all data sources involved in the MR study. [a] Data sources of exposure. [b] Data sources of outcome. GWAS ID, genome-wide association study identity; SNPs, single nucleotide polymorphisms; MR, mendelian randomization.

MR analysis

The MR analysis process is depicted in Fig. 1. We employed the inverse-variance weighted (IVW) as the primary analysis method, supplemented by MR-Egger[18] and weighted median[19] as additional analysis methods. The IVW method yields precise and reliable estimations under the assumption that all SNPs are effective IVs. The weighted median provides a consistent estimate when at least 50% of the weights originate from valid IVs. MR-Egger, allowing some SNPs to affect the outcome other than through exposure, furnishes a robust estimate of the causal effect in line with the IVW, albeit under a weaker assumption and with low efficiency.

For sensitivity analysis, Cochran's Q test was performed to detect data heterogeneity, where $p > 0.05$ implies no heterogeneity. MR-Egger intercept test was conducted to assess horizontal pleiotropy, with $p > 0.05$ indicating no pleiotropy. Additionally, a leave-one-out analysis was implemented to identify whether a single SNP drove the causal association by systematically excluding each SNP from the analysis.

Results

Selection of IVs

Following rigorous quality control procedures and the removal of palindromic SNPs, 111 SNPs associated with nine viruses were chosen as IVs. Specifically, 57 SNPs were associated with herpesviruses (including 7 SNPs for anti-EBV VCA IgG levels, 10 SNPs for anti-CMV IgG seropositivity, 13 SNPs for HHV-6 IE1A antibody levels, 8 SNPs for varicella and 19 SNPs for herpesviral infections), 39 SNPs were associated with respiratory viruses (including 31 SNPs for SARS-CoV-2 and 8 SNPs for anti-influenza A virus IgG levels), 10 SNPs were related to HIV disease of retrovirus, and 5 SNPs were for anti-HBsAg IgG levels. Supplementary Tables S1 summarize the main characteristics of each cohort. However, a portion of the identified SNPs did not meet the conventional threshold of $p < 5 \times 10^{-8}$ for MR studies. Consequently, a more lenient threshold of $p < 1 \times 10^{-5}$ and an LD threshold of $r^2 < 0.001$ were applied.
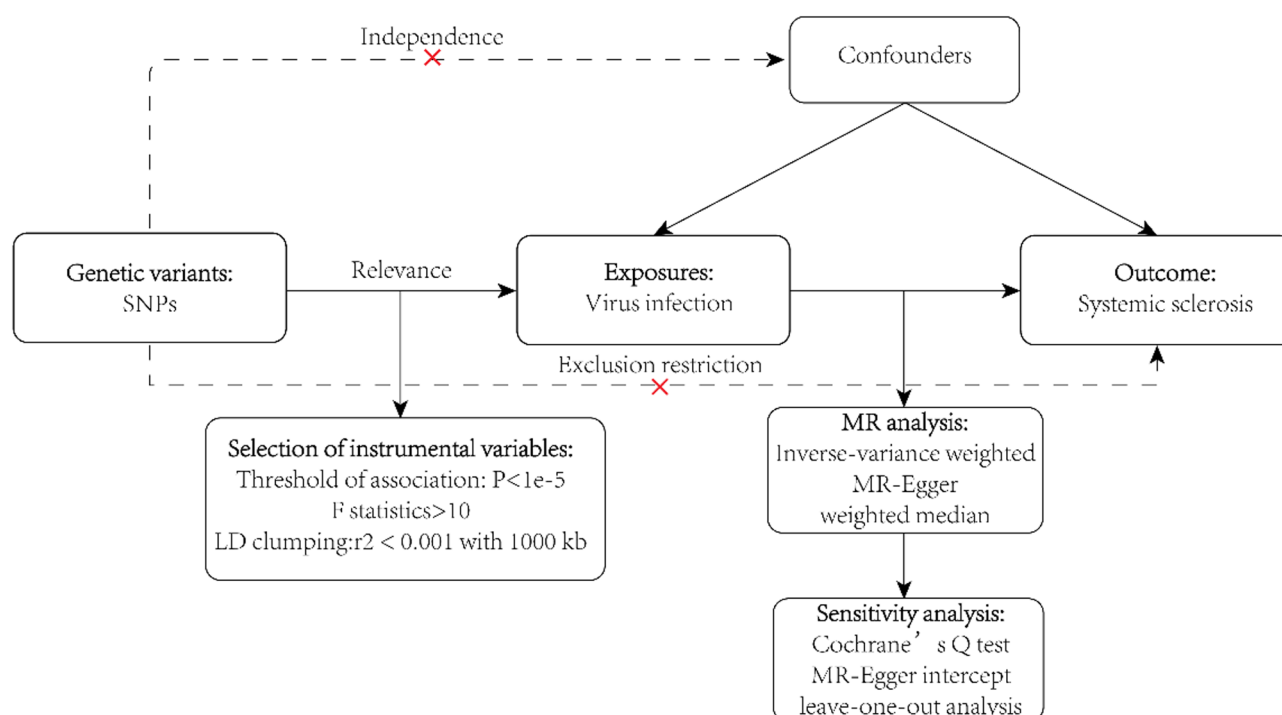
MR analysis

Detailed statistical results of the MR analysis are presented in Fig. 2.

Herpesviruses and SSc

Herpesviruses are enveloped viruses with double-stranded DNA genomes. Among them, EBV, CMV, HHV-6, VZV, and HSV are prevalent infections associated with humans.

In the preliminary investigation, higher genetically predicted anti-EBV VCA IgG levels were found to be associated with an increased risk of SSc. The IVW showed that each unit increase in anti-EBV VCA IgG levels elevated SSc risk by 3.4-fold (OR = 3.400, 95% CI = 1.093–10.571, $p = 0.035$). However, the null findings were confirmed in the estimates using MR-Egger (OR = 196.206, 95% CI = 2.468E-05-1.56E+09, $p = 0.544$) and the weighted median method (OR = 1.905, 95% CI = 0.430–8.443, $p = 0.396$). (Figures 2 and 3A)

In addition to anti-EBV VCA IgG levels, the associations of the other four targets with SSc were also explored. The IVW model concluded that there were no significant associations, including anti-CMV IgG seropositivity

Fig. 1. A flow diagram of the process in the MR analysis.
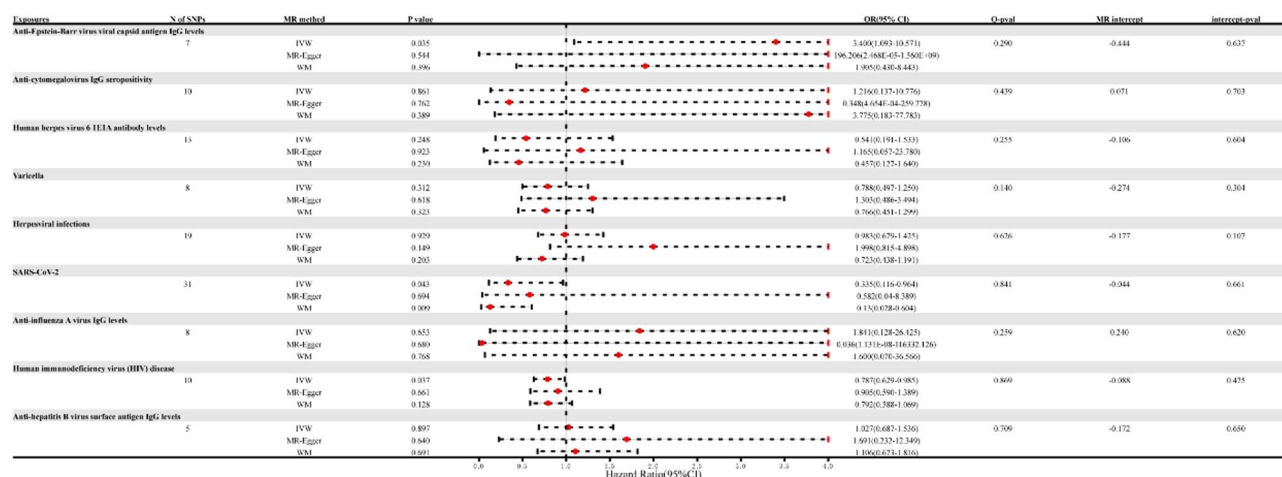
Fig. 2. Forest plots of MR estimates for the association between each virus infection and systemic sclerosis.

(OR = 1.216, 95% CI = 0.137–10.776, p = 0.861), HHV-6 IE1A antibody levels (OR = 0.541, 95% CI = 0.191–1.533, p = 0.248), varicella (OR = 0.788, 95% CI = 0.497–1.250, p = 0.312), and herpesviral infections (OR = 0.983, 95% CI = 0.679–1.425, p = 0.929). These results also did not demonstrate substantial changes when using MR-Egger and the weighted median methods (all p > 0.05). (Figures 2 and 3B-E)

Respiratory viruses and SSc
Respiratory viruses are a group of viruses that affect the respiratory tract, including influenza viruses (influenza A virus) and coronaviruses (SARS-CoV-2). There was a higher genetically predicted SARS-CoV-2 associated with a reduced risk of SSc in IVW (OR = 0.335, 95% CI = 0.116–0.964, p = 0.043). The weighted median analysis validated this result (OR = 0.130, 95% CI = 0.028–0.604, p = 0.009), enhancing confidence in the causal association. The MR-Egger analysis failed to detect a statistically significant association but indicated a similar changing trend (OR = 0.582, 95% CI = 0.040–8.398, p = 0.694). (Figures 2 and 3F) However, no significant causal associations were found between anti-influenza A virus IgG levels and SSc in any of the models (IVW: OR = 1.841, 95% CI = 0.128–26.425, p = 0.653). (Figures 2 and 3G)

Retrovirus and SSc
HIV is a retrovirus that causes acquired immunodeficiency syndrome. Evidence of a causal relationship between HIV disease (OR = 0.787, 95% CI = 0.629–0.985, p = 0.037) and SSc was observed, with HIV being a protective factor for SSc. Although the results of MR-Egger (p = 0.661) and the weighted median (p = 0.128) were not statistically significant, the effect direction was the same as the IVW model (OR < 1). (Figures 2 and 3H)

Hepatitis virus and SSc
Hepatitis viruses, such as hepatitis B virus, specifically infect hepatocytes and trigger viral infections of the liver. Through MR analysis of the relationship between anti-hepatitis B virus surface antigen IgG levels and SSc, negative results were obtained that IVW (OR = 1.027, 95% CI = 0.687–1.536, p = 0.897), weighted median (OR = 1.106, 95% CI = 0.673–1.816, p = 0.691) and MR-Egger analysis (OR = 1.691, 95% CI = 0.232–12.349, p = 0.640) suggested the absence of a causal relationship. (Figures 2 and 3I)

Sensitivity analysis
The Cochran's Q statistic did not reveal evidence of heterogeneity between different genetic tools (all p > 0.05). Visual inspection of the funnel plots showed symmetrical distributions, indicating minimal bias in these causal associations. Furthermore, no horizontal pleiotropy was detected (all p > 0.05), underscoring the robustness of the MR analysis. Based on the results of the leave-one-out analysis, it was evident that no single genetic variant significantly influenced causation in this study. (Figures 4 and 5)

Discussion
Numerous viral infections have been proposed as potential triggering factors for SSc, but no study has definitively established causality. Our MR study revealed that anti-EBV VCA IgG levels increased the risk of SSc, while SARS-CoV-2 and HIV disease were associated with a reduced risk. However, anti-CMV IgG seropositivity, anti-HBsAg IgG levels, anti-influenza A virus IgG levels, HHV-6 IE1A antibody levels, varicella, and herpesviral infections were not found to be causally associated with the risk of SSc.

Considerable evidence suggested that virus infection played an essential role in the development of SSc. Observational studies indicated the levels of antibodies against EBV and HCMV antigens increased significantly in SSc patients compared with healthy controls[5,20–22]. Another study found significantly elevated levels of antibodies against different EBV antigens compared to antibodies targeting SSc-specific antigens in patients with SSc[23]. Additionally, EBV DNA, viral transcripts, and proteins have been detected in the skin of SSc patients[21].
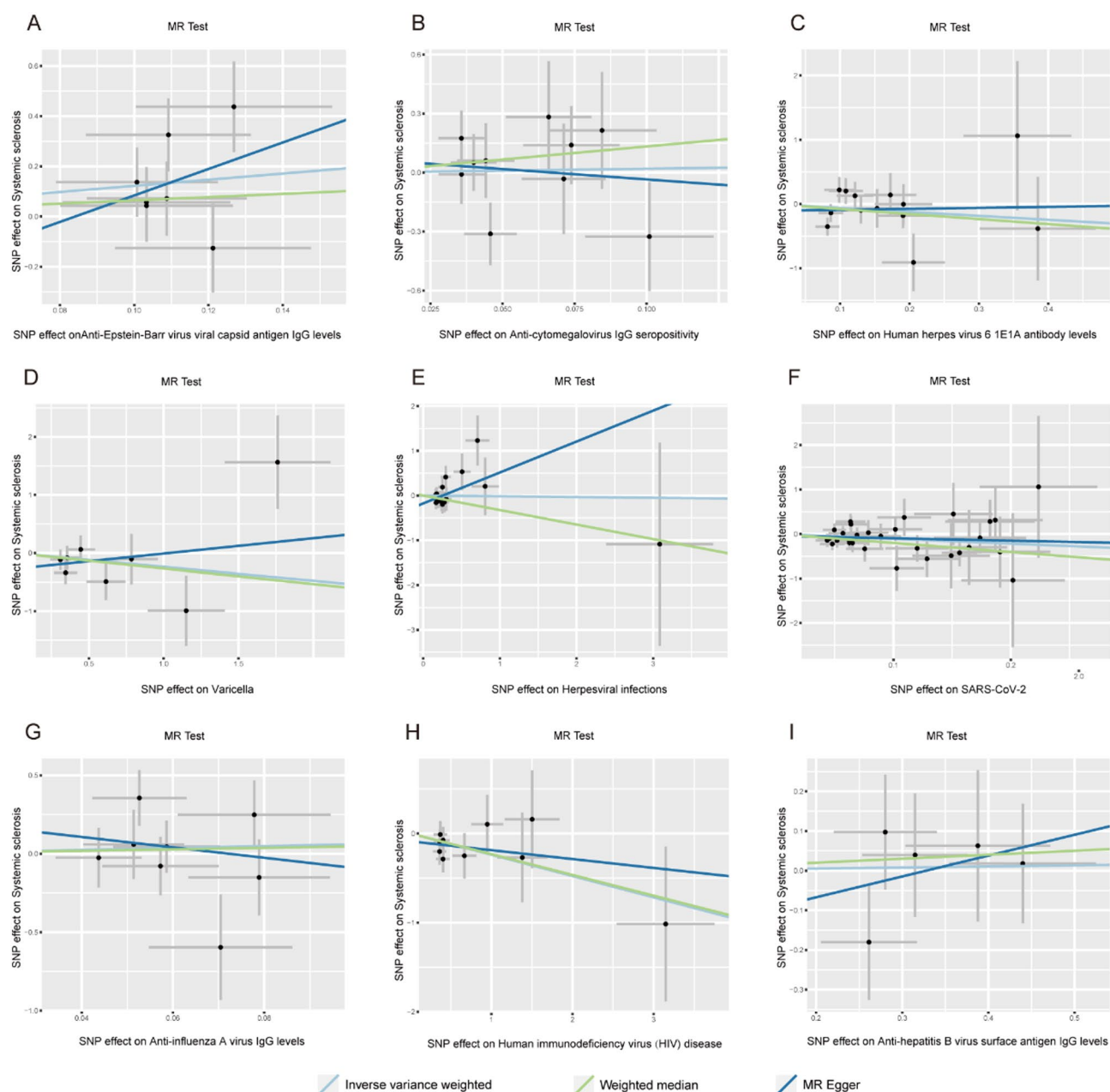
Fig. 3. Scatter plots for the effect of viral infections on systemic sclerosis.

Skin biopsies have demonstrated a higher prevalence of HHV-6 DNA in blood and peripheral tissue in SSc patients compared to controls[24]. Despite these observations, these findings did not provide direct evidence of causality. Our MR analysis offered more reliable evidence concerning the causal relationship between virus infection and SSc.

The SNPs selected in this study are not only associated with viral infections, but some sites have also been reported to have potential regulatory effects in other autoimmune diseases. For example, SNPs significantly associated with EBV VCA IgG levels (such as loci located in the HLA-DRB1 region) have also been identified as risk loci in GWAS studies of systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA)[25,26]. This phenomenon suggests that EBV may affect the occurrence of various autoimmune diseases through shared immune regulatory pathways, such as antigen presentation or interferon signaling. However, MR Egger regression did not detect significant level pleiotropy ($p > 0.05$), indicating that the effects of these SNPs on SSc are mainly mediated by EBV infection rather than directly acting on the immune pathway. It is worth noting that SNPs associated with HIV protective associations (such as rs333 in the CCR5 gene region) have been proven to reduce the risk of HIV infection and show pleiotropy in other infectious diseases (such as West Nile virus infection)[27]. These SNPs may indirectly affect the risk of SSc by modulating the broad-spectrum immune response of the host, but their specific mechanisms need to be further validated through functional research.
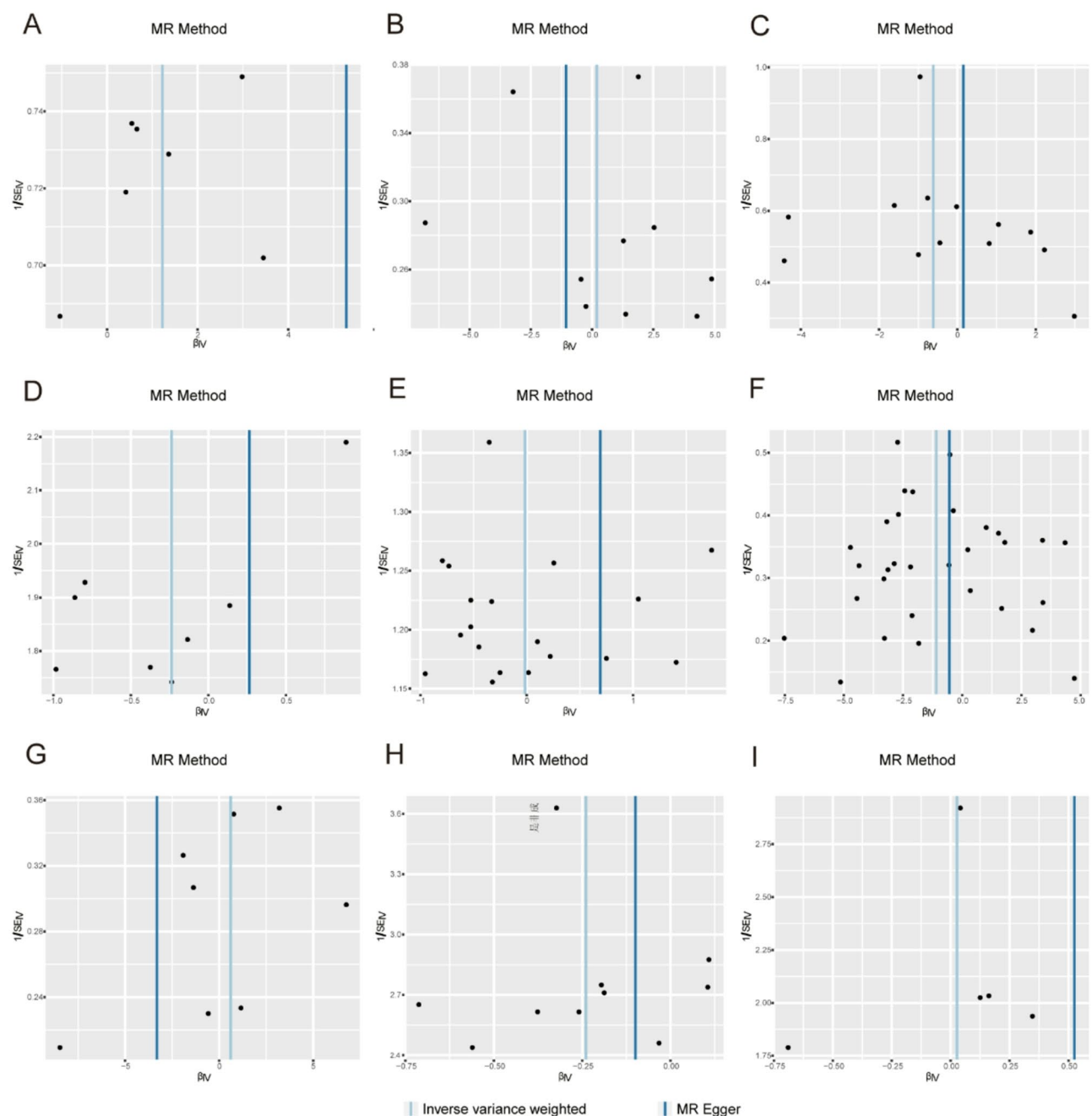
Fig. 4. Funnel plots of viral infections against the genetic associations with systemic sclerosis.

Previous studies have shown that EBV infection evidently plays a significant role in the etiology of many connective tissue diseases, including rheumatoid arthritis, systemic lupus erythematosus and Sjögren's syndrome[28]. This research indicated the causal relationship that high anti-EBV VCA IgG levels increased the risk of exacerbation of SSc. EBV, mostly in a latent form, infects the majority population worldwide with many immune evasion mechanisms and the ability to switch between latent and lytic reactivation stages[28]. High anti-EBV VCA IgG levels represent increased EBV viral load and the activation of antiviral mechanism and immune response. Though the precise mechanism by which EBV increases the risk of SSc has not been fully comprehended, several studies have proposed possible explanations. EBV can infect both fibroblasts and endothelial cells, the main SSc target cells, and activate abnormal innate immune responses[21]. Aberrant activation of EBV-Toll-like receptor (TLR) induces high expression of selected IFN- and transforming growth factorβ- (TGFβ-) responsive genes, such as CXCL9, OAS2, PAI-1and ET-1, promoting the fibrotic phenotype of fibroblasts in SSc[21,29]. More convincingly, TLR activation is unable to induce TGF response without EBV infection[21]. Aforementioned elucidation, linking EBV to the essential characteristics of SSc, including endothelial cell damage and abnormal fibroblast activation, may account for our causality of EBV on SSc[30].
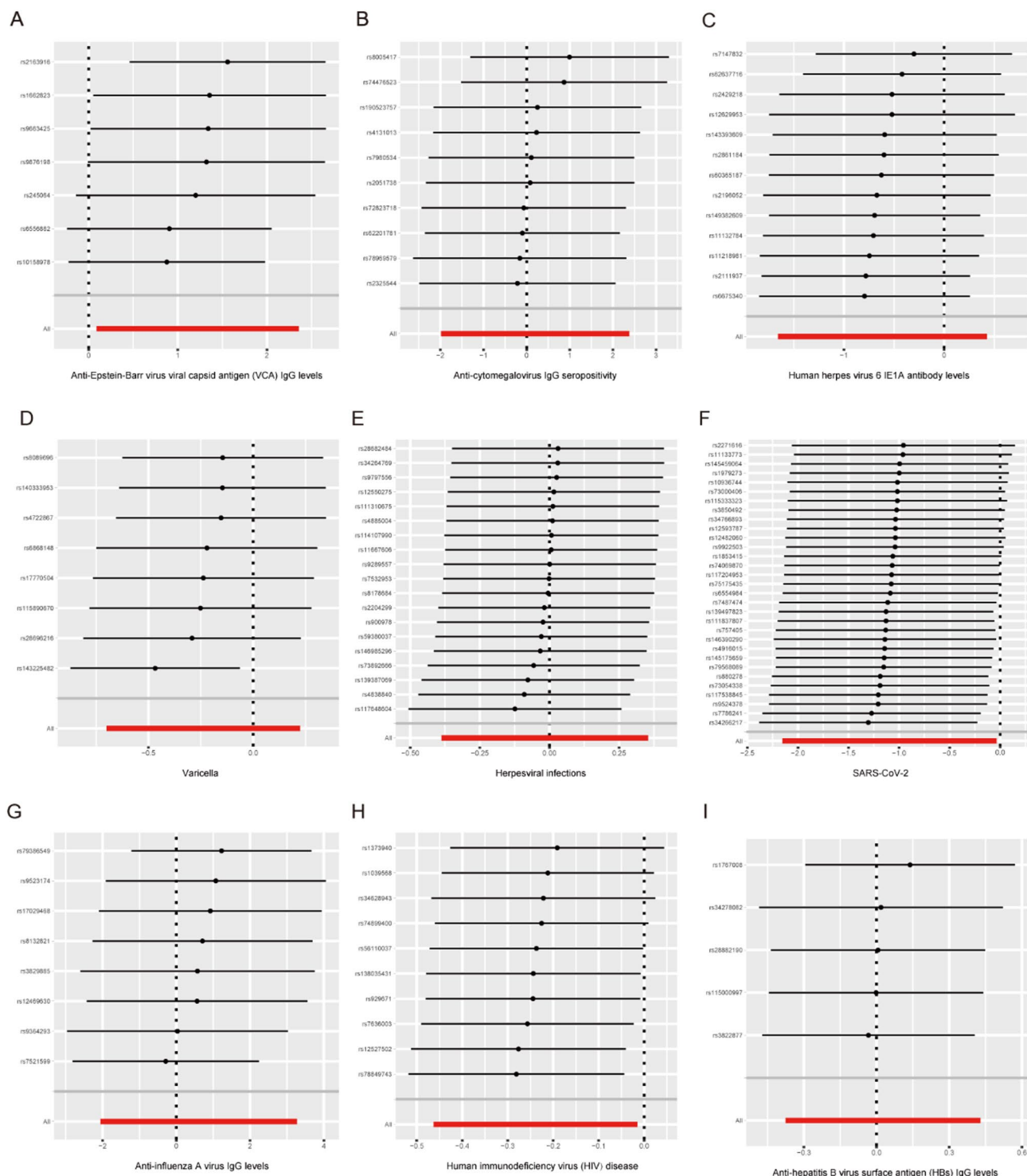
Fig. 5. Leave-one-out analysis plots for viral infections on systemic sclerosis.

Observational epidemiological studies have reported an increased risk of SSc associated with HIV and SARS-CoV-2. Recent studies indicated that patients developed features of SSc after HIV infection[31,32]. COVID-19, caused by SARS-CoV-2 can affect multiple organs and important functional units[33]. Serena Fineschi reported a case of SSc developing after COVID-19, suggesting a strong likelihood that SARS-CoV-2 infection may increase the risk of developing SSc[34]. A retrospective cohort study showed that patients with a positive PCR test result for SARS-CoV-2 had significantly higher risks of SSc[35]. However, our results conflict with some of these findings. What calls for special attention is that potential confounding factors cannot be overlooked in the study. On the one hand, environmental issues of drug action may influence the risk of SSc[36]. Most acquired immune deficiency syndrome (AIDS) patients developed SSc after treatment with antiretroviral drugs[31,32]. This may be

attributed to the treatment's ability to effectively mitigate the deleterious impact of HIV on the immune system. HIV binds to CD4 molecules on the surface of target cells through the outer membrane glycoprotein gp120 and invades T lymphocytes, monocytes/macrophages and dendritic cells, especially reducing the number of CD4 + T lymphocytes[37]. It not only compromises the body's antiviral immune defenses but also appreciably diminishes the potential for developing autoimmune disorders[38]. On the other hand, inherent limitations may have contributed to these contradictory results. For instance, clinical manifestations, immune responses, and pathogenic mechanisms similar between COVID-19 and SSc can affect the diagnosis results[39]. On the other hand, inherent limitations may have contributed to these conflicting results. Firstly, the risk of false negative classification, where some infected individuals may be misclassified as the control group due to weak antibody response or short duration (such as asymptomatic/mild patients or delayed sampling time)[40]. Secondly, GWAS data did not differentiate between viral strains (such as Alpha and Delta variants) and the severity of infection, and the immune escape ability of different strains or long-term immune activation in critically ill patients may affect the risk of SSc[41]. Thirdly, the positive correlation between SARS-CoV-2 and SSc reported in other studies may be due to stricter case confirmation in clinical cohorts (such as high titer antibodies or specific organ involvement), or inclusion of patients with newly developed autoimmune phenomena after COVID-19[35]. The MR analysis in this study is based on genetic prediction of lifelong infection risk, rather than short-term effects after acute infection, which may explain the differences compared to other observational studies. In the future, it is necessary to combine virology (such as strain specific antibodies) and clinical stratification data (such as infection severity) to further validate the robustness of the results. Therefore, more intensive and comprehensive research is necessary to elucidate the connection between them. Notably, a cohort study has demonstrated that patients with SSc are at a significantly increased risk of developing severe COVID-19 and fatal outcomes compared to the general population[42]. Therefore, we recommend that patients with unclear symptoms after the COVID-19 outbreak should be promptly investigated and tested for autoantibodies.

Our MR analysis indicated no genetic causality between anti-CMV IgG seropositivity and the risk of SSc, conflicting with some previous studies. Reportedly, a 33-year-old woman developed SSc after an acute HCMV infection[43]. Amino acid sequence homologies between HCMV and human proteins suggest that HCMV induces SSc through molecular mimicry mechanism and antibody cross-reaction[44]. UL94, an HCMV-derived protein, exhibits homology with NAG-2(tetraspan novel antigen-2) which is highly expressed in endothelial cells[45]. The cross-reaction between anti-UL94 HCMV antibodies and NAG-2 leads to endothelial damage and fibrogenesis[46]. However, the level of anti-UL94 antibodies in SSc patients is similar to that in healthy individuals, and virus response is only for a short peptidyl sequence, not the entire antigen[46]. Although anti-UL57 antibody responses are more frequent in SSc patients than controls, they also occur in other autoimmune diseases and there is no cross-reaction between Anti-UL57 antibodies and SSc-specific antigens[46]. However, the complex mechanisms have not been understood based on the current evidence. These characteristics may explain our null result regarding the causality of HCMV on SSc in this study.

HHV-6 has been considered as one of the possible etiological factors of SSc. An observational study detected higher HHV-6 DNA levels in blood and peripheral tissue in SSc patients than in controls. The presence of HHV-6 A in skin biopsies, in line with HHV-6 tropism for endothelial cells, was also reported[24]. Endothelial progenitor cells with impaired angiogenic function and the regression of capillaries and small vessels were detected in the peripheral blood of SSc patients which were suspected of being caused by HHV-6[47]. Unexpectedly, our study did not find a causal relation between HHV-6 and the increased risk of SSc. Several explanations may elucidate our results. Evidence suggested an increased level of tolerogenic human leucocyte antigen (HLA-G) in SSc with HHV-6 infection compared to HHV-6 negative SSc and the highest titer of anti-U94 antibodies in SSc with both HHV-6 and high HLA-G, indicating HHV-6 impaired angiogenetic ability through U-94 mediated HLA-G production[24,47]. The impairment of immune response against HHV-6 is extremely evident in SSc expressing the killer immunoglobulin-like receptors (KIR2DL2), which suggests HHV-6 alone does not increase the risk of SSc, when together with KIR2DL2, elevates SSc risk[24,47].

To the best of our knowledge, our study is the first to relatively comprehensively expound the causal relations between various virus infections and SSc. Several strengths of our MR study contribute to its robustness. First, strict selection conditions of SNPs strengthen exposure-outcome associations and mitigate the interference of confounding factors. Second, reverse causation is ruled out since the genotype is determined at conception. Third, employing three different models to estimate the causality, coupled with sensitivity test, supports the stability and accuracy of the results. At the same time, it is undeniable that there are limitations in our study. First, it should be noted that the identification of SSc cases in the GWAS database relies on medical records, which may introduce two potential biases. Diagnosis bias may occur as confirmed SSc cases are more likely to be moderate-to-severe patients, while early or limited cases might be overlooked. Detection bias may arise because patients who are serologically positive, such as those with anti-Scl-70 antibody positivity, are more likely to be diagnosed and included in the database. Despite these potential biases, the case definition method based on strict clinical criteria can maximize the accuracy of disease phenotyping, which is a clear advantage compared to studies based on patient-reported data. Secondly, the small sample size may result in low power to detect causality. Furthermore, we adopted the significance threshold of $p < 1 \times 10^{-5}$ rather than the standard $p < 5 \times 10^{-8}$, which may introduce bias due to weak instrumental variables, despite calculating and using F-statistic to define strong instrumentals. Although this threshold has been used in previous studies, caution is warranted in interpreting results[48,49]. Additionally, due to the reliance on medical system records, there may be an underestimation of mild or undiagnosed SSc patients. What's more, SSc exhibits higher morbidity in women but has higher mortality in men[50]. Unfortunately, due to the statistical limitations, we cannot complement stratified MR analysis based on gender. Finally, to minimize potential heterogeneity, the genetic background was limited to European ancestry. Therefore, caution should be exercised when generalizing results to other races.

Conclusions

Our MR study suggested a causal association of EBV infection on the elevated risk of SSc and the causal effects of HIV and SARS-CoV-2 on the decreased risk of SSc, providing valuable insights into the genetic underpinning of associations between virus infections and SSc. These findings may benefit the prevention, diagnosis, and prognosis of SSc. Despite its strengths, the study's limitations underscore the need for further research to refine our understanding of the complex interactions between viral infections and SSc.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Hughes, M. & Herrick, A. L. Systemic sclerosis. *Br. J. Hosp. Med. (Lond).* **80** (9), 530–536 (2019).
2. Stochmal, A., Czuwara, J., Trojanowska, M. & Rudnicka, L. Antinuclear antibodies in systemic sclerosis: an update. *Clin. Rev. Allergy Immunol.* **58** (1), 40–51 (2020).
3. Randone, S. B., Guiducci, S. & Cerinic, M. M. Systemic sclerosis and infections. *Autoimmun. Rev.* **8** (1), 36–40 (2008).
4. Smatti, M. K. et al. Viruses and autoimmunity: A review on the potential interaction and molecular mechanisms. *Viruses* ;**11**(8). (2019).
5. Arnson, Y. et al. The role of infections in the immunopathogensis of systemic sclerosis–evidence from serological studies. *Ann. N Y Acad. Sci.* **1173**, 627–632 (2009).
6. Machhua, S. et al. Detection of Epstein-Barr virus in systemic sclerosis patients: A molecular and serological based study. *Int. J. Rheum. Dis.* **25** (12), 1431–1436 (2022).
7. Nystad, T. W. & Myrmel, H. Prevalence of primary versus reactivated Epstein-Barr virus infection in patients with VCA IgG-, VCA IgM- and EBNA-1-antibodies and suspected infectious mononucleosis. *J. Clin. Virol.* **38** (4), 292–297 (2007).
8. Svicher, V. et al. New markers in monitoring the reactivation of hepatitis B virus infection in immunocompromised hosts. *Viruses* ;**11**(9). (2019).
9. Yu, X., Lophatananon, A., Mekli, K., Muir, K. R. & Guo, H. Exploring the causal role of the immune response to varicella-zoster virus on multiple traits: a phenome-wide Mendelian randomization study. *BMC Med.* **21** (1), 143 (2023).
10. Emdin, C. A., Khera, A. V., Kathiresan, S. & Mendelian Randomization *JAMA* ;**318**(19):1925–1926. (2017).
11. Peng, H. et al. Association between systemic sclerosis and risk of lung cancer: results from a pool of cohort studies and Mendelian randomization analysis. *Autoimmun. Rev.* **19** (10), 102633 (2020).
12. Villanueva-Martin, G. et al. The effect of body fat distribution on systemic sclerosis. *J. Clin. Med.* ;**11**(20). (2022).
13. Yan, W., Luo, Q., Nie, Q., Wang, H. & Wu, J. Association between systemic sclerosis and left ventricle dysfunction: findings from observational studies. *Heliyon* **9** (3), e14110 (2023).
14. Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* **375**, n2233 (2021).
15. Scepanovic, P. et al. Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med.* **10** (1), 59 (2018).
16. Initiative, C-H-G. The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* **28** (6), 715–718 (2020).
17. Butler-Laporte, G. et al. Genetic determinants of Antibody-Mediated immune responses to infectious diseases agents: A Genome-Wide and HLA association study. *Open. Forum Infect. Dis.* **7** (11), ofaa450 (2020).
18. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32** (5), 377–389 (2017).
19. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40** (4), 304–314 (2016).
20. Sternbaek, L. et al. Increased antibody levels to stage-specific Epstein-Barr virus antigens in systemic autoimmune diseases reveal a common pathology. *Scand. J. Clin. Lab. Invest.* **79** (1–2), 7–16 (2019).
21. Farina, A. et al. Epstein-Barr virus infection induces aberrant TLR activation pathway and fibroblast-myofibroblast conversion in scleroderma. *J. Invest. Dermatol.* **134** (4), 954–964 (2014).
22. Neidhart, M. et al. Increased serum levels of antibodies against human cytomegalovirus and prevalence of autoantibodies in systemic sclerosis. *Arthritis Rheum.* **42** (2), 389–392 (1999).
23. Fattal, I. et al. Epstein-Barr virus antibodies mark systemic lupus erythematosus and scleroderma patients negative for anti-DNA. *Immunology* **141** (2), 276–285 (2014).
24. Caselli, E. et al. HHV-6A infection and systemic sclerosis: clues of a possible association. *Microorganisms* ;**8**(1). (2019).
25. Bentham, J. et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47** (12), 1457–1464 (2015).
26. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506** (7488), 376–381 (2014).
27. Lim, J. K. et al. Genetic deficiency of chemokine receptor CCR5 is a strong risk factor for symptomatic West nile virus infection: a meta-analysis of 4 cohorts in the US epidemic. *J. Infect. Dis.* **197** (2), 262–265 (2008).
28. Houen, G. & Trier, N. H. Epstein-Barr virus and systemic autoimmune diseases. *Front. Immunol.* **11**, 587380 (2020).
29. Farina, A. & Farina, G. A. Fresh insights into disease etiology and the role of microbial pathogens. *Curr. Rheumatol. Rep.* **18** (1), 1 (2016).
30. Gilbane, A. J., Denton, C. P. & Holmes, A. M. Scleroderma pathogenesis: a pivotal role for fibroblasts as effector cells. *Arthritis Res. Ther.* **15** (3), 215 (2013).
31. Hasan, S., Aqil, M. & Panigrahi, R. HIV-Associated systemic sclerosis: literature review and a rare case report. *Int. J. Environ. Res. Public. Health* ;**19**(16). (2022).
32. Okong'o, L. O., Webb, K. & Scott, C. HIV-associated juvenile systemic sclerosis: a case report. *Semin Arthritis Rheum.* **44** (4), 411–416 (2015).
33. Hu, B., Guo, H., Zhou, P. & Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19** (3), 141–154 (2021).
34. Fineschi, S. Case report: systemic sclerosis after Covid-19 infection. *Front. Immunol.* **12**, 686699 (2021).
35. Chang, R. et al. Risk of autoimmune diseases in patients with COVID-19: A retrospective cohort study. *EClinicalMedicine* **56**, 101783 (2023).
36. Abbot, S., Bossingham, D., Proudman, S., de Costa, C. & Ho-Huynh, A. Risk factors for the development of systemic sclerosis: a systematic review of the literature. *Rheumatol. Adv. Pract.* **2** (2), rky041 (2018).

37. Fanales-Belasio, E., Raimondo, M., Suligoi, B. & Butto, S. HIV virology and pathogenetic mechanisms of infection: a brief overview. *Ann. Ist Super Sanita.* **46** (1), 5–14 (2010).
38. Moir, S. & Fauci, A. S. Insights into B cells and HIV-specific B-cell responses in HIV-infected individuals. *Immunol. Rev.* **254** (1), 207–224 (2013).
39. Denton, C. P. et al. COVID-19 and systemic sclerosis: rising to the challenge of a pandemic. *J. Scleroderma Relat. Disord.* **6** (1), 58–65 (2021).
40. Dan, J. M. et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* ;**371**(6529). (2021).
41. Planas, D. et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596** (7871), 276–280 (2021).
42. Hoffmann-Vold, A. M. et al. Systemic sclerosis in the time of COVID-19. *Lancet Rheumatol.* **4** (8), e566–e75 (2022).
43. Shin, J. G. et al. Inhibitory effects of tricyclic antidepressants (TCAs) on human cytochrome P450 enzymes in vitro: mechanism of drug interaction between TCAs and phenytoin. *Drug Metab. Dispos.* **30** (10), 1102–1107 (2002).
44. Efthymiou, G. et al. A comprehensive analysis of antigen-specific antibody responses against human cytomegalovirus in patients with systemic sclerosis. *Clin. Immunol.* **207**, 87–96 (2019).
45. Lunardi, C., Bason, C., Corrocher, R. & Puccetti, A. Induction of endothelial cell damage by hCMV molecular mimicry. *Trends Immunol.* **26** (1), 19–24 (2005).
46. Marou, E. et al. Human cytomegalovirus (HCMV) UL44 and UL57 specific antibody responses in anti-HCMV-positive patients with systemic sclerosis. *Clin. Rheumatol.* **36** (4), 863–869 (2017).
47. Ferri, C. et al. Insights into the knowledge of complex diseases: environmental infectious/toxic agents as potential etiopathogenetic factors of systemic sclerosis. *J. Autoimmun.* **124**, 102727 (2021).
48. Chang, M. J. et al. Mendelian randomization analysis suggests no associations of herpes simplex virus infections with systemic lupus erythematosus. *J. Med. Virol.* **95** (3), e28649 (2023).
49. Li, P. et al. Association between gut microbiota and preeclampsia-eclampsia: a two-sample Mendelian randomization study. *BMC Med.* **20** (1), 443 (2022).
50. Hughes, M. et al. Gender-related differences in systemic sclerosis. *Autoimmun. Rev.* **19** (4), 102494 (2020).

Author contributions
T.C., S.-X. Z. and Y.-H. H. designed the study. H.-Y. Z., T.-R. W., and R.-L. L. collected and analyzed data. S.-X. Z., H.-Y. Z., and T. C. contributed to the analysis and interpretation of the data. T.-R. W., R.-L. L., and Z.-N. J. wrote the manuscript. All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. S.-X. Z. and Y.-H. H. had full access to all of the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Declarations

Competing interests
The authors declare no competing interests.

Ethics
The original GWAS data have obtained prior approval from relevant ethics review boards and the summary-level statistics don't contain any personal information, thus no additional ethics approval was required.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-08361-z.

Correspondence and requests for materials should be addressed to Y.-H.H. or S.-X.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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