

Mapping disparities in viral infection rates using highly multiplexed serology

Alejandra Piña,¹ Evan A. Elko,¹ Rachel Caballero,² Morgan Metrailler,¹ Mary Mulrow,² Dan Quan,^{2,3,4} Lora Nordstrom,² John A. Altin,⁵ Jason T. Ladner¹

AUTHOR AFFILIATIONS See affiliation list on p. 21.

ABSTRACT Despite advancements in medical interventions, the disease burden caused by viral pathogens remains large and highly diverse. This burden includes the wide range of signs and symptoms associated with active viral replication as well as a variety of clinical sequelae of infection. Moreover, there is growing evidence supporting the existence of sex- and ethnicity-based health disparities linked to viral infections and their associated diseases. Despite several well-documented disparities in viral infection rates, our current understanding of virus-associated health disparities remains incomplete. This knowledge gap can be attributed, in part, to limitations of the most commonly used viral detection methodologies, which lack the breadth needed to characterize exposures across the entire virome. Additionally, virus-related health disparities are dynamic and often differ considerably through space and time. In this study, we utilize PepSeq, an approach for highly multiplexed serology, to broadly assess an individual's history of viral exposures, and we demonstrate the effectiveness of this approach for detecting infection disparities through a pilot study of 400 adults aged 30–60 in Phoenix, AZ. Using a human virome PepSeq library, we observed expected seroprevalence rates for several common viruses and detected both expected and previously undocumented differences in inferred rates of infection between our male/female and Hispanic/non-Hispanic White individuals.

IMPORTANCE Our understanding of population-level virus infection rates and associated health disparities is incomplete. In part, this is because of the high diversity of human-infecting viruses and the limited breadth and sensitivity of traditional approaches for detecting infection events. Here, we demonstrate the potential for modern, highly multiplexed antibody detection methods to greatly increase our understanding of disparities in rates of infection across subpopulations (e.g., different sexes or ethnic groups). The use of antibodies as biomarkers allows us to detect evidence of past infections over an extended period, and our approach for highly multiplexed serology (PepSeq) allows us to measure antibody responses against hundreds of viruses in an efficient and cost-effective manner.

KEYWORDS highly multiplexed serology, virome, health disparities, antibody repertoire, PepSeq, virus, salivirus A, hepatitis A virus, *Enterovirus C*, human herpesviruses, human immunodeficiency virus, human adenovirus D

Despite advancements in the development of medical interventions, the global burden of disease caused by viral pathogens remains substantial and highly diverse. This burden includes a wide range of morbidities associated with active viral replication ranging in severity from fever, muscle aches, and rash to encephalitis, immunosuppression, respiratory failure, and congenital birth defects (1–4). Additionally, it includes an array of clinical sequelae of infection (e.g., Guillain-Barre syndrome,

Editor Genevieve G. Fouda, Duke Human Vaccine Institute, Durham, North Carolina, USA

Address correspondence to Jason T. Ladner, Jason.Ladner@nau.edu.

Alejandra Piña and Evan A. Elko contributed equally to this article. Author order was determined alphabetically by first name.

The authors declare no conflict of interest.

See the funding table on p. 21.

Received 19 February 2024

Accepted 21 July 2024

Published 20 August 2024

Copyright © 2024 Piña et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

multisystem inflammatory syndrome, long coronavirus disease), many of which remain poorly understood (5–7). Viral infections have even been linked to the onset of a number of non-communicable diseases, such as myocarditis (8), diabetes (9), celiac disease (10), obesity (11), multiple sclerosis (12, 13), cancer (14), and Alzheimer's disease (15).

Although the health effects caused by viral infections are observed widely in the general population, currently documented national trends highlight several sex- and ethnicity-based health disparities in the prevalence of viral infections and thus, virus-associated disease. For example, in the United States (US), seroprevalence of human papillomavirus (HPV) and herpes simplex virus 2 (HSV-2) have been shown to be roughly twice as high among women compared to men (16, 17). In contrast, human immunodeficiency virus 1 (HIV-1) disproportionately affects men in the US, particularly men who have sex with men. In 2021, 69% of new HIV-1 diagnoses in the US were among men who have sex with men, despite this group representing only 3.9% of the US population (18, 19). Ethnic disparities have also been documented for HIV-1 in the US. A 2015–2019 CDC surveillance report showed the incidence of HIV-1 infection among Hispanics was four times higher than that among non-Hispanic Whites. Furthermore, disparities in viral infections can vary in space and time. For example, over 12 years the average annual incidence of hepatitis A virus (HAV) infection in a Native American population decreased from 289 to 6 per 100,000 people due to vaccination efforts (20). Vector-borne viruses provide striking examples of geographical disparities. For example, the dengue virus is transmitted by mosquitoes of the genus *Aedes* that live primarily in tropical regions (21). In 2020, the number of locally transmitted cases of dengue virus was 0.025 per 100,000 people in the continental United States, compared to 23.68 cases per 100,000 people in Puerto Rico (22).

Despite these well-documented disparities in viral infection rates, our understanding of virus-associated health disparities remains incomplete. In part, this is because the most commonly used methodologies for detecting viral infections are limited in their breadth, both in terms of the number of viruses they can detect and/or the period during which detection is possible. Molecular assays detect viral nucleic acids and therefore lack sensitivity in cases where the infection has already been cleared or if the sampled fluid does not contain the virus (23). In contrast, serological assays detect antiviral antibodies that can persist for years after exposure due to the body's long-lived humoral immune response. However, the most commonly used serological assays, such as the enzyme-linked immunosorbent assay (ELISA), only test for one virus (and typically one protein) at a time (24).

Recent advances in serological methods have overcome previous limitations in breadth and are enabling unprecedented views into the viral exposure histories of individuals (25–29). Using current approaches for highly multiplexed serology (e.g., PepSeq [30] and PhiP-Seq [31]) it is now possible to characterize antibody binding to 100,000s of antigens in a single assay using <1 µL of blood. In this study, we utilize the PepSeq platform to demonstrate the potential of highly multiplexed serology to broadly characterize differences in seroprevalence across the human virome between different demographic groups. Through the characterization of antiviral antibodies in samples collected over ~1 month within a single healthcare system in Phoenix, AZ, we document significant differences in infection rates between the Hispanic White (HW) and non-Hispanic White (NHW) populations for several viruses, including some that are rarely included in population-level surveys.

RESULTS

Study population

Between late May and early June 2020, we collected 400 remnant serum samples from 11 Valleywise Health facilities, a large safety net hospital system in Phoenix, AZ. These samples were distributed equally among four subpopulations: 100 HW men, 100 HW women, 100 NHW men, and 100 NHW women. To minimize the impact of age-related

differences in seropositivity, we limited our focal age range to 30–60 years, and the age distributions were not significantly different (*t*-test) between genders or ethnicities, with mean age ranging between 45.1 and 46.2 (Fig. 1A). We also investigated payor source for each sampled individual (Fig. 1B), as this can serve as a proxy for socioeconomic status (32). Overall, the vast majority of the individuals included in our study were either covered by a government insurance plan (52.8%; Tricare, Medicare, and/or Medicaid) or were uninsured (31.3%; self-pay). Only ~13.1% of the individuals in our study were covered by commercial insurance plans and this percentage did not vary considerably among our subpopulations, though we observed a higher rate of commercial payor for NHW females (20.2%) compared to the other three groups (10%–11.2%). However, we did observe substantial differences among our subpopulations in the proportions covered by either government programs or uninsured. The NHW populations were covered in higher proportions by government insurance programs (HW-M: 44.9%, HW-F: 16%, NHW-M: 79.8%, NHW-F: 72.7%), while the HW populations were more likely to be uninsured (HW-M: 39.8%, HW-F: 72%, NHW-M: 10.1%, NHW-F: 3%).

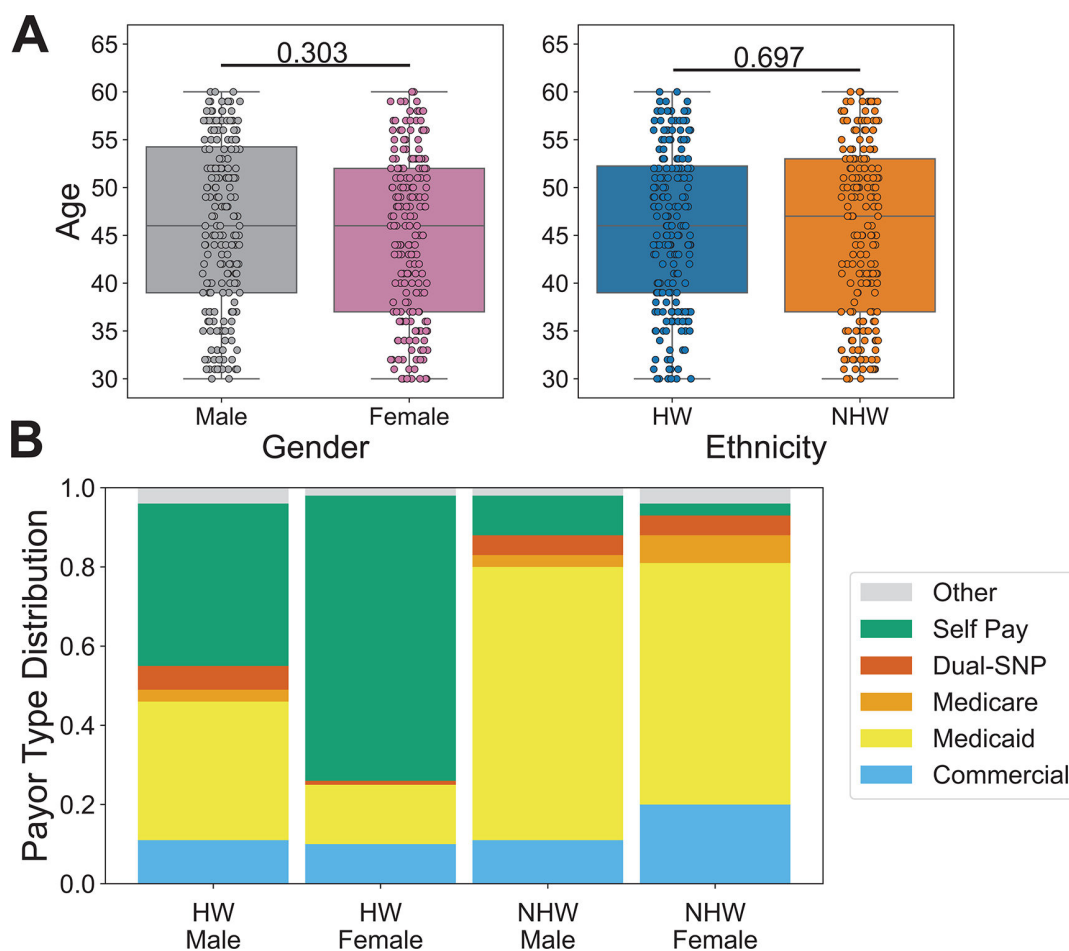


FIG 1 Demographics of the study population. Remnant serum samples were collected from 400 individuals at Valleywise Health in Phoenix, AZ, and assayed using the HV1 PepSeq library. (A) The study population ranged in age from 30 to 60 years old and there was no statistically significant difference in the age distributions between genders and ethnicities (*t*-tests). Individual *t*-test *P*-values comparing male/female and HW/NHW ages are indicated above the respective plots. Each circle represents an individual. The line within each box represents the median, while the lower and upper bounds of each box represent the first and third quartiles, respectively. The whiskers extend to points that lie within 1.5 interquartile ranges of the first and third quartiles. (B) Payor sources varied substantially among focal subpopulations. Payor source was consolidated into six general categories (Table S4) and is shown for each subpopulation. Dual-SNP refers to any dual special needs plans for individuals who qualify for both Medicare and Medicaid.

PepSeq analysis

All 400 serum samples were assayed, in duplicate, using our human virome version 1 (HV1) PepSeq library (26), and we obtained an average of 2.2M Illumina sequencing reads per sample, which equates to an average of 9.2 reads per unique HV1 peptide per sample. Ten samples were excluded from further analysis due to having low raw read counts (<488,000 raw sequence reads) or a lower than normal correlation between replicates (Z score Pearson's correlation <0.6), which may indicate the occurrence of molecular bottlenecks or contamination of one of the replicates while performing the assay (30). Therefore, all the analyses presented here include a total sample size of 390. Among the excluded samples were five HW males, two NHW males, one HW female, and two NHW females.

Based on visual comparisons of experimental samples and buffer-only negative controls (Fig. S1), as well as the analysis of a separate group of negative controls that were not considered in the formation of bins or normalization of the data, we chose a set of four different Z score thresholds (10, 15, 20, and 25) for identifying enriched peptides. Higher thresholds are expected to have reduced sensitivity, but increased specificity. To estimate the false positive rate at each of these thresholds, we analyzed all pairwise combinations of 9 buffer-only negative controls ($n = 36$). We observed an average of 5.5, 1.1, 0.28, and 0.17 putatively enriched peptides from these control pseudoreplicate analyses for thresholds of 10, 15, 20, and 25, respectively. In contrast, from the assays of serum samples, we observed an average of 1,338, 1,074, 929, and 830 enriched peptides for thresholds of 10, 15, 20, and 25, respectively. This equates to expected false positive rates of approximately 0.41%, 0.1%, 0.03%, and 0.02%, respectively.

To broadly characterize enrichment patterns within our data set, we averaged the number of enriched peptides across all Z score thresholds for each sample. We did not observe significant differences between the average number of enriched peptides by gender (mean values: F = 1,012.79, M = 1,074.14; t -test P -value = 0.133) or ethnicity (mean values: HW = 1,050.13, NHW = 1,036.24; t -test P -value = 0.734) (Fig. 2A). We also observed no significant correlation between the number of enriched peptides and age (Fig. 2B; Pearson correlation P -value = 0.176).

Estimating seropositivity from PepSeq

We first converted our lists of enriched peptides into putative virus species-level serostatus calls for each individual using the deconv module of PepSIRF with fixed seropositivity score thresholds ranging from 20 to 600. We then compared our PepSeq estimates of seroprevalence with published estimates based on more traditional singleplex assays (Fig. S2). This analysis included 18 virus species with published seroprevalence studies in the United States (Table S1) (33–55). There was a significant positive correlation between the PepSeq estimates of seroprevalence and the seroprevalence estimates found in the literature, and this correlation was strongest when using a score threshold of 40 for determining seropositivity (P -value = 2.02×10^{-3} , Pearson $R = 0.677$) (Fig. 3A; Fig. S2). This score threshold requires ≥ 2 peptides in support of each seropositivity call. However, despite the overall correlation, we observed several viruses with substantially higher or lower-than-expected seroprevalence estimates (Fig. 3A; Fig. S2).

To assess the impact of seropositivity score thresholds on the sensitivity and specificity of our PepSeq assay for individual virus species, we generated receiver operating characteristic (ROC) curves with seropositivity score thresholds ranging from 0 to 14,000. This analysis focused on five viruses, including two that are generally well-documented in medical records (HIV-1 and hepatitis C virus [HCV]) and three with commercially available ELISAs that are expected to have intermediate seroprevalence in the US: cytomegalovirus (CMV), herpes simplex virus 1 (HSV-1), and HSV-2 (33, 37, 43–46). In general, PepSeq determinations of seropositivity for these viruses were in strong agreement with expectations based on medical records and the ELISA assays (Fig. 3B and C). Average area under the curve (AUC) values (across our four Z score thresholds)

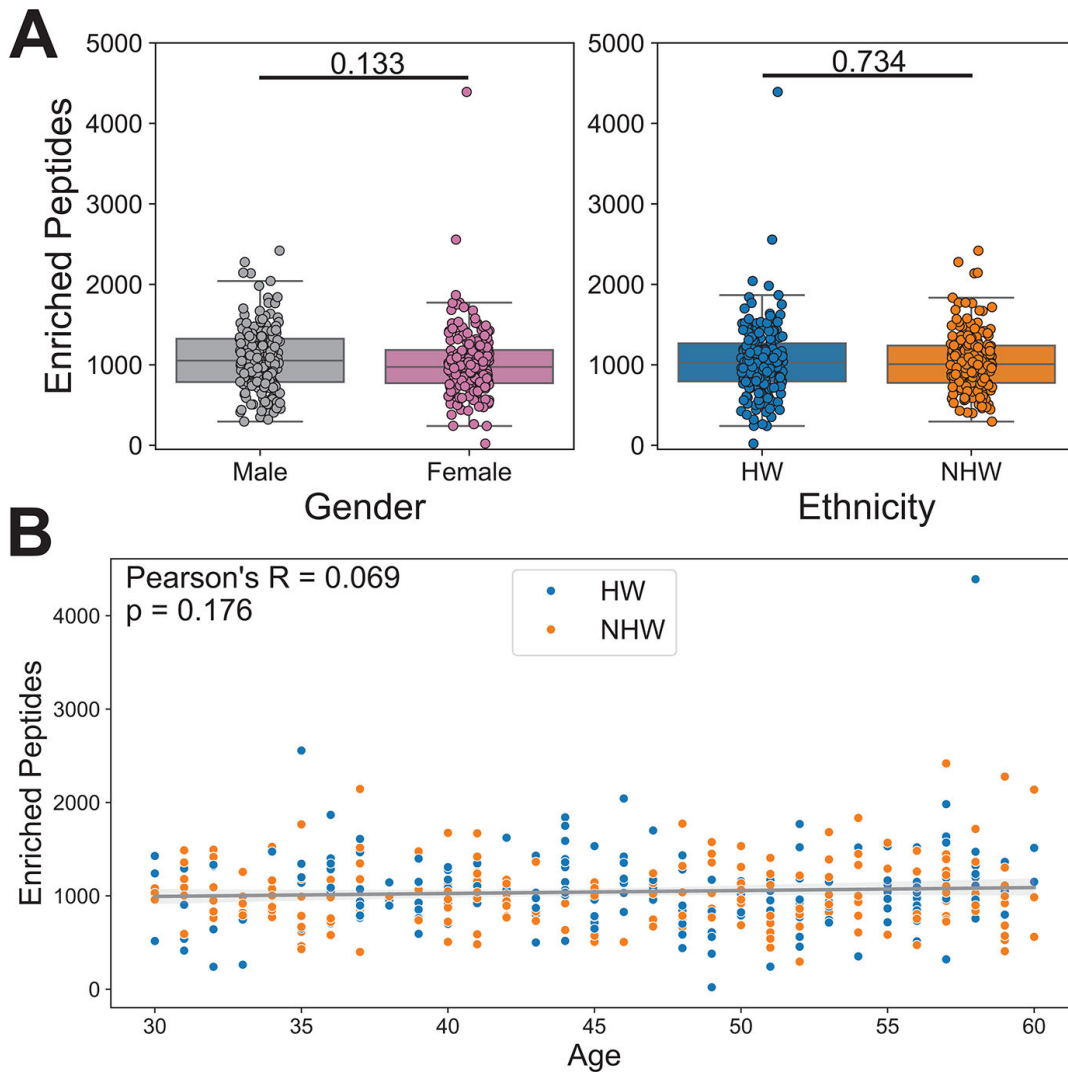


FIG 2 PepSeq identifies similar overall levels of antibody reactivity against viral peptides between genders and ethnicities. (A) Box plots depicting the average number of enriched PepSeq peptides for each sample across four Z score thresholds ($Z = 10, 15, 20,$ and 25). Average number of enriched peptides: female = 1,012.79, male = 1,074.14, HW = 1,050.13, NHW = 1,036.24. Individual *t*-test *P*-values comparing male/female and HW/NHW enriched peptide counts are indicated above the respective plots. Each circle represents an individual. The line within each box represents the median, while the lower and upper bounds of each box represent the first and third quartiles, respectively. The whiskers extend to points that lie within 1.5 interquartile ranges of the first and third quartiles. (B) Scatter plot with a best-fit line showing an average number of enriched peptides by age. Ethnicity is indicated by the color of the points, HW = blue and NHW = orange. The gray diagonal line indicates the best-fit linear regression with the shaded gray areas showing the 95% confidence interval. Pearson correlation was used to test for significance (*P*-value = 0.176).

were 0.94 and 0.88 for HIV-1 and HCV, respectively (Fig. 3B), when compared to each individual's documented infection status. Average AUC scores were 0.99, 0.99, and 0.88 for CMV, HSV-1, and HSV-2, respectively (Fig. 3C), when comparing to the results of singleplex ELISA assays run on a subset of our focal population ($n = 78-87$, Table S2). The optimal seropositivity score thresholds (maximum of sensitivity + specificity) for these five viruses ranged from 60 to 600 (HSV-2 = 60, HSV-1 = 200, HIV-1 = 200, HCV = 200, CMV = 600).

To explore the impact of representation bias within our PepSeq library on deviations from expected seroprevalence, we compared the maximum possible seropositivity score from our PepSeq assay to the difference between our seroprevalence estimate (using a fixed score threshold of 40) and expected seroprevalence from the literature. Overall, we observed a strong positive correlation between these values (Pearson correlation

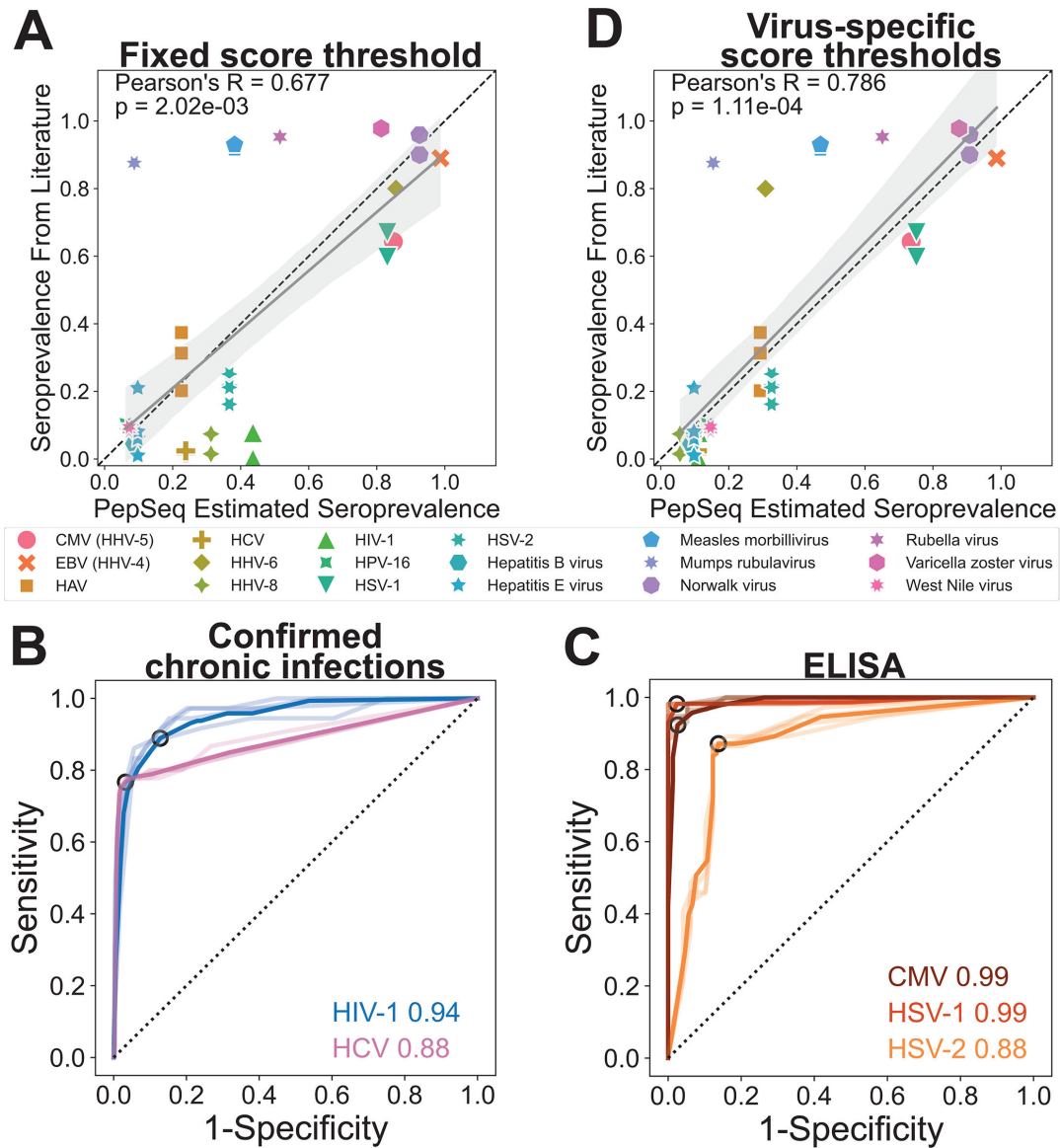


FIG 3 PepSeq-based determinations of seropositivity correlate with published estimates of seroprevalence, independent singleplex ELISA assays, and known chronic infection status. (A) PepSeq estimated seroprevalence across the full data set, using a Z score threshold of 15 and a seropositivity score threshold of 40 for all viruses, compared to published seroprevalence values from studies in the United States (Table S1). The gray diagonal line indicates the best-fit linear regression with the shaded gray areas showing the 95% confidence interval. The dashed line indicates $x = y$. Pearson's R-value and P-value are shown in the top left. (B, C) Receiver operating characteristic (ROC) curves showing the performance of PepSeq seropositivity calls across a range of score thresholds compared to confirmed chronic infections (B) and commercially available ELISAs (C). Lines for each of the four Z score thresholds are shown in lighter colors and the bold color depicts the average of the values for the four Z scores. The point corresponding to the optimal seropositivity score threshold is circled in black. The numbers on the right indicate the area under the curve (AUC) for the averaged lines. (D) PepSeq estimated seroprevalence across the full data set, using virus-specific, representation-normalized seropositivity score thresholds, compared to published seroprevalence values from studies in the United States; displayed as described for panel A. Abbreviations: CMV, cytomegalovirus; EBV, Epstein Barr virus; HCV, hepatitis C virus; HHV, human herpesvirus.

P -value = $9.12e-05$, Fig. S3A). This indicated that, when using a fixed seropositivity score threshold, our approach had a tendency to overestimate seroprevalence for viruses that were overrepresented in our library and to underestimate seroprevalence for viruses that were underrepresented in our library. To better account for representation bias in our PepSeq library, we utilized distinct seropositivity score thresholds for each virus. We set these score thresholds using each virus' maximum possible seropositivity score (a proxy for representation in the PepSeq library) and an inferred relationship between

this score and the optimal score thresholds estimated using singleplex ELISAs with an enforced plateau at a seropositivity score of 20 for the least represented viruses (equivalent to ~1 enriched peptide supporting a seropositivity call; Fig. S3B). Seroprevalence estimates using these representation-normalized seropositivity score thresholds exhibited an even stronger correlation with published estimates of seroprevalence, with no viruses exhibiting unusually high estimates of seroprevalence for this population as a whole (Pearson correlation P -value = 1.11×10^{-4} , Fig. 3D).

Using the representation-normalized seropositivity score thresholds, we averaged the number of seropositive species per sample across all Z score thresholds to broadly characterize patterns within our subpopulations, and we found no significant differences (t -test) in the average number of predicted seropositive virus species by gender (mean values: F = 37.64, M = 38.78) or by ethnicity (mean values: HW = 38.84, NHW = 37.57) (Fig. 4A). However, we did observe a significant positive correlation between age and the number of predicted seropositive virus species (Pearson correlation P -value = 0.0004, Fig. 4B).

Identification of disparities

To identify statistically significant differences in estimated seroprevalence among our subpopulations, we fit a binomial generalized linear model (GLM) with a single dependent variable (serostatus) and three independent variables (ethnicity, gender, and age). For this analysis, we chose to use our representation-normalized seropositivity score thresholds, which provided the best correlation to published estimates of seroprevalence across 18 different viruses (Fig. 3D; Fig. S2). However, we also ran all of the same analyses using a fixed seropositivity score threshold of 40 for all viruses, and the results were very similar (not shown).

In general, across virus species, we observed higher rates of seropositivity in older individuals, but there was no consistent directional change associated with gender or ethnicity (Fig. S4). However, no individual viruses exhibited significant correlations between age and serostatus after correcting for multiple tests, which may be related to the limited range of ages included in this study (30–60 years old). We also found no significant differences in estimated serostatus between genders after multiple test corrections, but three viruses exhibited significant differences prior to multiple test corrections that were consistent across all Z score thresholds (Fig. 5A). Two viruses, HIV-1 and HCV, exhibited higher seroprevalence in males (consistent with medical records for study participants) and one virus, Sapporo virus, exhibited higher seroprevalence in females (Fig. 5A).

We also observed several viruses with significant correlations between serostatus and ethnicity, and these patterns were generally consistent across our four Z score thresholds (Fig. 5B). In total, nine virus species exhibited ethnicity P -values < 0.05 across all four Z score thresholds: CMV (a.k.a. human herpesvirus 5), HSV-1, HSV-2, human herpesvirus 7 (HHV-7), HAV, salivirus A (SaV-A), parechovirus A (PeV-A), *Enterovirus C* (EV-C), and human adenovirus D (HAdV-D). After Bonferroni correction for multiple tests, seven of these remained significant at ≥ 1 Z score threshold and six remained significant across all four Z score thresholds (Fig. 5B). Within the HW subpopulation, we observed significantly higher seroprevalence for CMV, HSV-1, HAdV-D, HAV, and SaV-A across all four Z score thresholds (Fig. 5B). Within the NHW subpopulation, we observed significantly higher seropositivity for HHV-7 across all thresholds and for PeV-A at a Z threshold of 20.

For all nine viruses showing significant or nearly significant differences in seroprevalence between ethnicities, we also looked at the relationship between seropositivity and insurance status. Specifically, using a Z score threshold of 15, we compared seroprevalence estimates among three subsets of our study population: insured NHWs ($n = 183$), insured HWs ($n = 86$), and uninsured HWs ($n = 108$). Uninsured NHWs were excluded because of a low sample size ($n = 13$) but generally showed seroprevalence estimates similar to insured NHWs. Our results showed that for all nine viruses, estimated seroprevalence within insured HWs was intermediate between the NHW-insured and

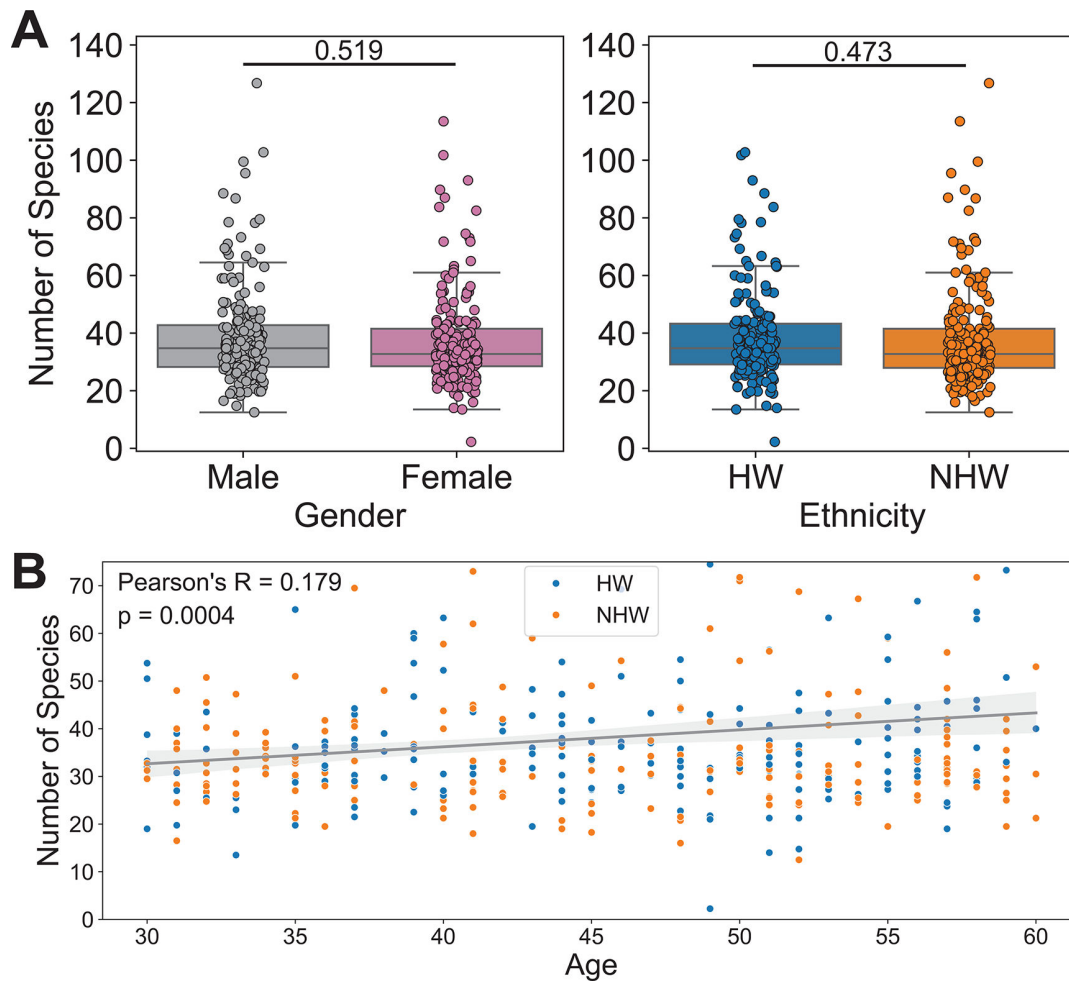


FIG 4 PepSeq-based estimates of seropositivity correlate with age. (A) Box plots depicting the number of virus species selected as seropositive by the PepSIRF deconv algorithm (with representation-normalized seropositivity score thresholds) for each sample, divided according to gender (left) and ethnicity (right). Average number of putatively seropositive virus species: female = 37.64, male = 38.78, HW = 38.84, NHW = 37.57. *t*-Tests comparing male/female and HW/NHW were non-significant (*P*-values indicated above the respective plots). Each circle represents an individual. The line within each box represents the median, while the lower and upper bounds of each box represent the first and third quartiles, respectively. The whiskers extend to points that lie within 1.5 interquartile ranges of the first and third quartiles. (B) Scatter plot with a best-fit line showing an average number of predicted seropositive virus species by age. Ethnicity is indicated by the color of the points, HW = blue and NHW = orange. The gray diagonal line indicates the best-fit linear regression with the shaded gray areas showing the 95% confidence interval. Pearson correlation was used to test for significance (*P*-value = 0.0004).

HW-uninsured subpopulations (Fig. 5C). When considering all nine viruses together, we observed a significant difference between normalized estimates of seroprevalence in the insured HW group compared to the uninsured HW group (absolute values, paired *t*-test *P*-value = 0.0002). For all six viruses with higher seroprevalence among HWs, seroprevalence was highest among uninsured individuals, and for all three HWs, seroprevalence was lowest in the HW uninsured group.

Vaccines are available for two of the virus species with higher seroprevalence in our HW population (HAV and EV-C, which includes poliovirus), and as a consequence, the differences we observed could be due to differences in rates of either natural infection or vaccination. Therefore, we further dissected the antibody responses against these viruses by examining protein- and peptide-level reactivity profiles. Notably, all HAV vaccines approved for use in the US are inactivated and it has been shown that antibody responses to natural infection and vaccination can be differentiated by measuring the response to non-structural proteins, to which a response will only be generated with a natural infection (i.e., when there is virus replication and production of non-structural

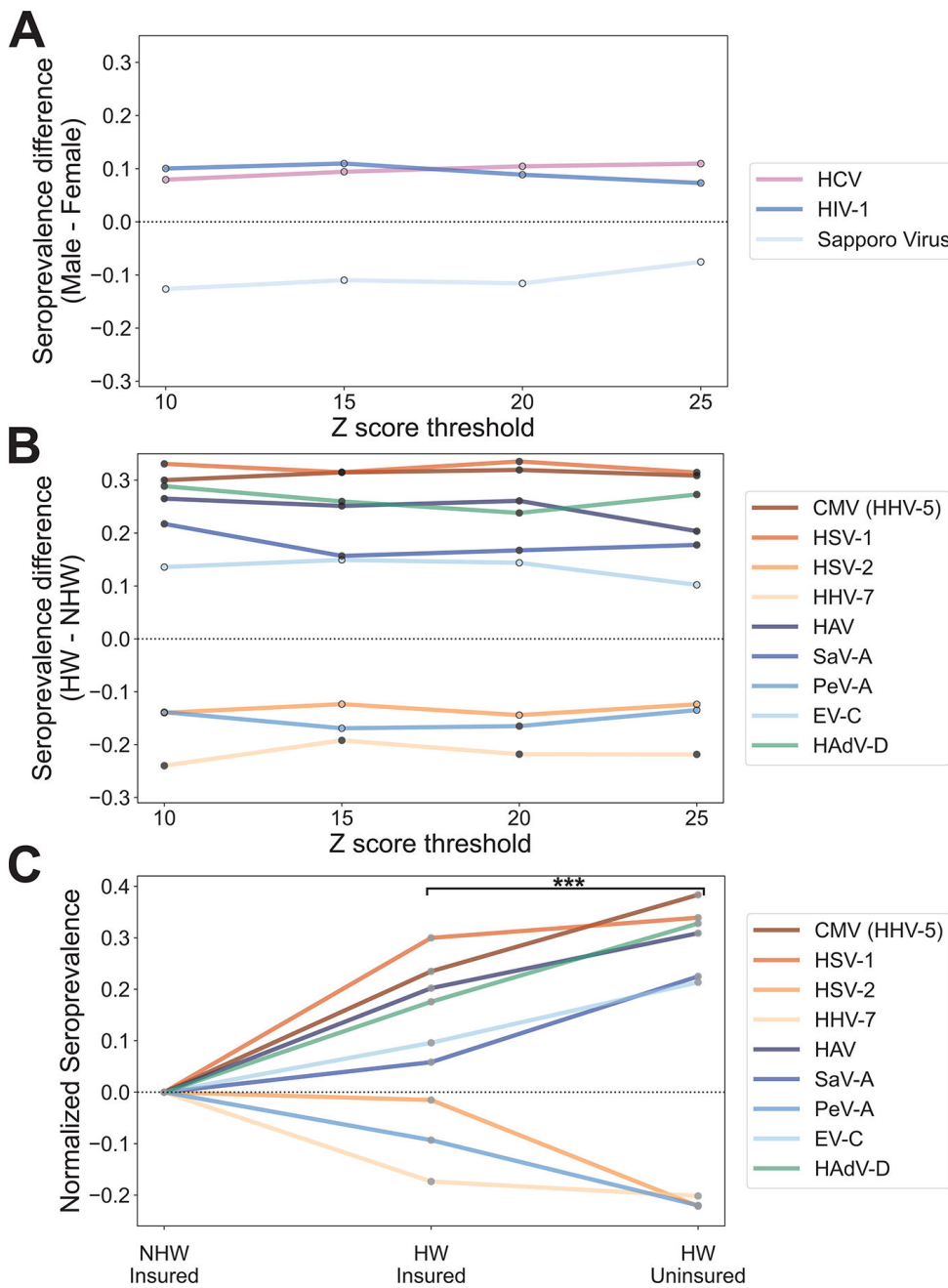


FIG 5 Significant differences in seroprevalence by ethnicity and payor status. (A, B) Line plots depicting virus species with significant differences (P -value < 0.05) in seropositivity between (A) males and females or (B) HWs and NHWs calculated by fitting a generalized linear model at each Z score threshold before (outlined points) and after (filled points) Bonferroni correction for multiple tests. Negative differences indicate higher seroprevalence in females or NHWs and positive differences indicate higher seroprevalence in males or HWs for panels A and B, respectively. (C) Line plot depicting normalized seroprevalence for the same nine viruses shown in panel B, with values calculated separately for insured and uninsured individuals. Seroprevalence is being shown for a Z score threshold of 15 and was normalized against the value for insured NHWs. The asterisk indicates the significant increase in the absolute value of the normalized seroprevalence for uninsured HWs compared to insured HWs across all nine viruses (paired t -test P -value = 0.0002). Abbreviations: SaV-A, salivirus A; PeV-A, parechovirus A; EV-C, *Enterovirus C*; HAdV-D, human adenovirus D.

proteins) (56). Therefore, to examine the role of natural exposure to HAV in the observed difference in seropositivity, we mapped all enriched HAV peptides for each seropositive

sample across the HAV proteome. We did not observe a significant difference between HW and NHW cohorts in the proportion of seropositive individuals with ≥ 1 enriched peptide from an HAV non-structural protein (Fisher's exact test P -value = 0.288). We observed high rates of reactivity against non-structural HAV proteins across both groups (HW = 56/81, NHW = 26/33) (Fig. 6A). These findings indicate that most of the seropositive individuals in our study have likely been naturally infected by HAV and that the observed difference in seroprevalence between HW and NHW is probably not driven by differences in vaccination rate.

Next, we sought to determine the role of poliovirus vaccination on the increased seropositivity to EV-C in HWs compared to NHWs. In this case, both inactivated and attenuated (replication-competent) vaccines were used in the US prior to 2000, and both are still administered in Mexico, which borders Arizona and is the most common source of immigrants in the state (57). Therefore, instead of comparing protein-level patterns of antibody reactivity, we exploited patterns of amino acid divergence within species to compare reactivity across different strains of EV-C. Specifically, we reran our estimates of seropositivity after separating the three polioviruses (UniProt accessions can be found in Table S3) from the rest of EV-C. In this new analysis, we saw no significant difference in estimated serostatus for poliovirus (P -value = 0.28) with 139 (72%) HW and 150 (77%) NHW positive samples (Fig. 6B). However, we did observe a highly significant difference in seropositivity between HWs and NHWs for "Other EV-C" strains, with 130 (67%) and 66 (34%) seropositive samples, respectively (P -value < 0.0001) (Fig. 6B). These results indicate that the observed disparity in EV-C seropositivity between HWs and NHWs is not driven by differences in natural infection or vaccination with poliovirus but is likely caused by differences in infection rate with other EV-C viruses.

To determine whether there were particular non-polio EV-C viruses that were driving the disparity seen in the "Other EV-C" group, we further analyzed the reactive peptides assigned to this group. First, we split the 27 International Committee on Taxonomy of Viruses (ICTV) listed EV-C isolates into six phylogenetic clades (Fig. 7A, one clade includes only the polioviruses). Next, we assigned each "Other EV-C" peptide to the most similar EV-C clade (based on shared amino acid 7-mers) and assigned these peptide scores equivalent to the number of contained 7-mers that are unique to that clade. We then calculated relative peptide scores for each EV-C clade by summing the scores for (i) all "Other EV-C" peptides in the full HV1 PepSeq library (null distribution) and (ii) the subsets of enriched "Other EV-C" peptides observed in the HW and NHW samples that were seropositive for "Other EV-C" (Fig. 7B; Fig. S5). We observed antibody reactivity to peptides assigned to all five "Other EV-C" clades, and the clade-specific relative peptide scores varied substantially between individuals (Fig. S5). These results suggest that a variety of different EV-C viruses may be contributing to the observed disparity in EV-C seroprevalence between HWs and NHWs. However, some of these clades (e.g., EV-C_1–3) may be more common than others, based on differences in the relative peptide scores between the expected ("Full Library") and observed ("NHW," "HW") distributions (Fig. 7B).

DISCUSSION

In this study, we used PepSeq, a highly multiplexed serology platform, to broadly assess virus infection histories and identify differences in seropositivity among various subsets of the population served by Valleywise Health in Phoenix, AZ. PepSeq allows for 100,000s of peptide antigens to be simultaneously assayed for antibody reactivity; thus, it has the potential to facilitate the comprehensive characterization of differences in viral infection rates among subsets of a community. In contrast, the singleplex nature of traditional serological techniques (e.g., ELISA) has required that previous studies focus on a small number of high-priority viruses (58, 59).

Although we have previously shown that PepSeq-based measures of antibody reactivity correlate well with the results of singleplex ELISAs (27, 28), this is the first time we have used PepSeq to estimate seropositivity across the human virome using a cross-sectional cohort. In general, our results were consistent with expectations from

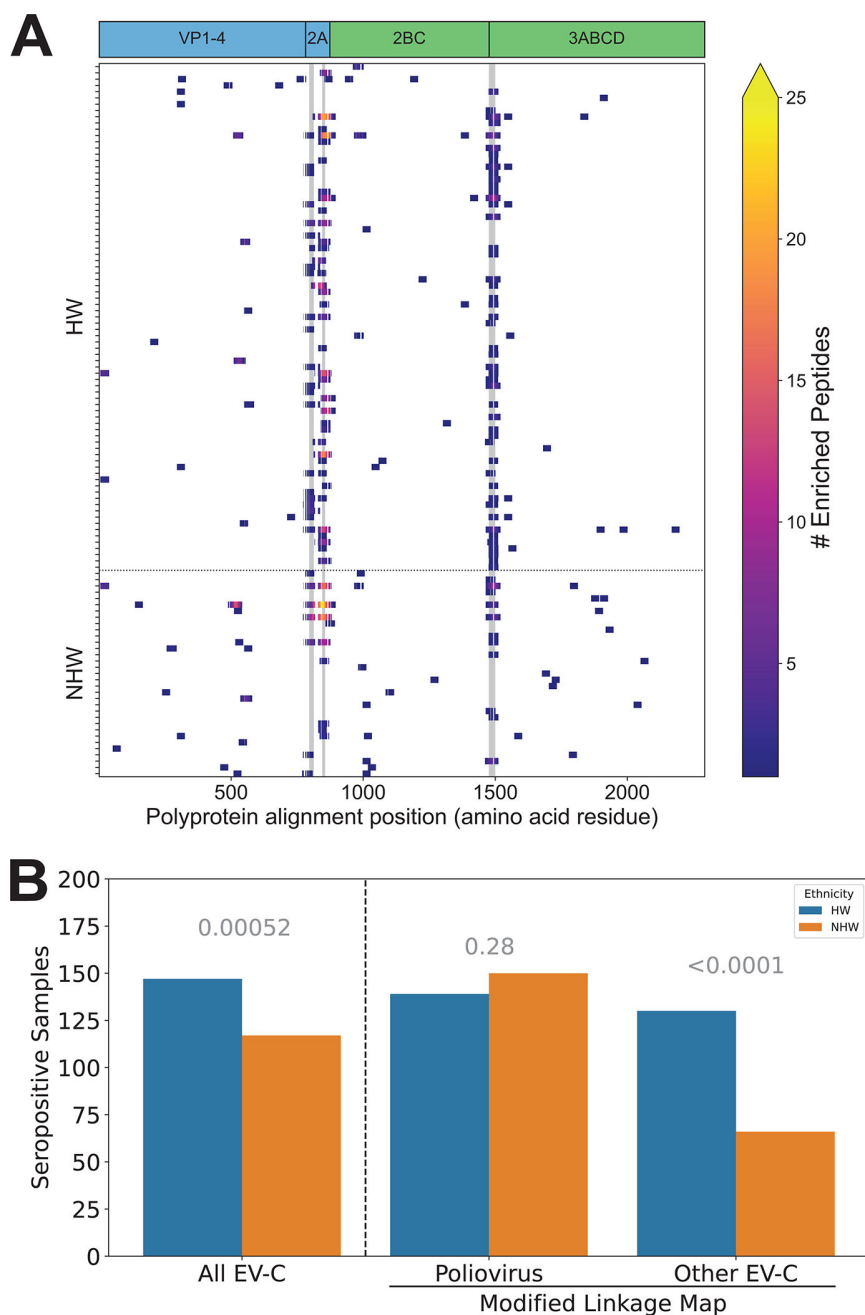


FIG 6 Ethnicity-based disparities in viral infection rates are not driven by vaccination against HAV or poliovirus. (A) Enriched peptides and public epitopes were identified in both structural and non-structural HAV proteins. Heatmap of enriched peptides across the HAV proteome for all samples seropositive for HAV with a Z score threshold of 15. Samples are broken up by ethnicity on the y-axis and position within the HAV polyprotein is shown on the x-axis (amino acid residues, alignment coordinates). The three most commonly reactive epitopes within these samples are highlighted with gray vertical markings. The positions of individual viral proteins are indicated across the top; blue = structural (vaccination or infection) and green = non-structural (infection only). (B) Bar plot showing the number of seropositive HW (total $n = 194$) and NHW (total $n = 196$) individuals. Seropositivity was calculated using the original linkage map with poliovirus peptides assigned to the EV-C virus species (“all EV-C”), or with a modified linkage map where EV-C were reassigned to one of two categories depending on whether they shared ≥ 1 7-mer with polioviruses (“Poliovirus”) or not (“Other EV-C”). P -values for binomial GLM comparing the impact of ethnicity (HW, NHW) on seropositivity are indicated in gray above each category.

previously published studies, particularly once we adjusted for representation bias in our PepSeq library (Fig. 3D). For example, we observed a general trend toward higher seropositivity with increasing age, a pattern that has been reported for a variety of viruses (33–35, 60). We observed this pattern at two levels: (i) a positive correlation between an individual’s age and the number of seropositive virus species called (Fig. 4B) and (ii) negative age-associated GLM coefficients for most viruses, indicative of higher seroprevalence with increased age (Fig. S4). We also found that overall estimates of seroprevalence for individual viruses were broadly consistent with expectations from both molecular and singleplex serological surveys. For example, with a Z score threshold of 15, seven viruses had estimated seropositivities $\geq 90\%$ (Fig. S6). Among these were five common respiratory viruses, for which the expected seroprevalence is near 100% by adulthood (61): human rhinoviruses A, B, and C, human orthopneumovirus (aka respiratory syncytial virus) (62), and human respirovirus 3 (a.k.a. human parainfluenza virus 3) (63). Also included in this set are human gammaherpesvirus 4 (a.k.a. Epstein-Barr virus) and Norwalk virus (a.k.a. norovirus), consistent with published serological surveys of adults in the US (34, 52). Additionally, we found that relative seroprevalence estimates for closely related viruses were generally consistent with documented differences in prevalence. For example, while we estimated a seroprevalence for the more common human respirovirus 3 of 94%, the less common “parainfluenza” viruses (human respirovirus 1 and rubulaviruses 2 and 4) were estimated to have seroprevalences of 35%–43% (64). Similarly, we estimated the seroprevalence of HSV-1 (75%) to be $\sim 2.2\times$ higher than that of HSV-2 (33%) (Fig. 3D), reflecting known differences in the prevalence of these related viruses (44).

However, even after accounting for representation bias, there were several viruses for which our PepSeq-based estimates of seroprevalence were substantially lower than expected (Fig. 3D). This included all three viruses that are part of the MMR live attenuated vaccine (measles morbillivirus, mumps rubulavirus, and rubella virus), as well as human herpesvirus 6 (HHV-6) (65). One important limitation of the PepSeq technology is that it is generally limited to measuring antibodies that bind to linear epitopes, as the peptide antigens cannot accurately represent epitopes that are dependent on tertiary and quaternary structures. If the antibody responses against these viruses are directed primarily against conformational epitopes, this could explain the lower sensitivity in our highly multiplexed assay. Future PepSeq libraries may be able to mitigate these virus-specific reductions in sensitivity by including peptides designed to mimic known

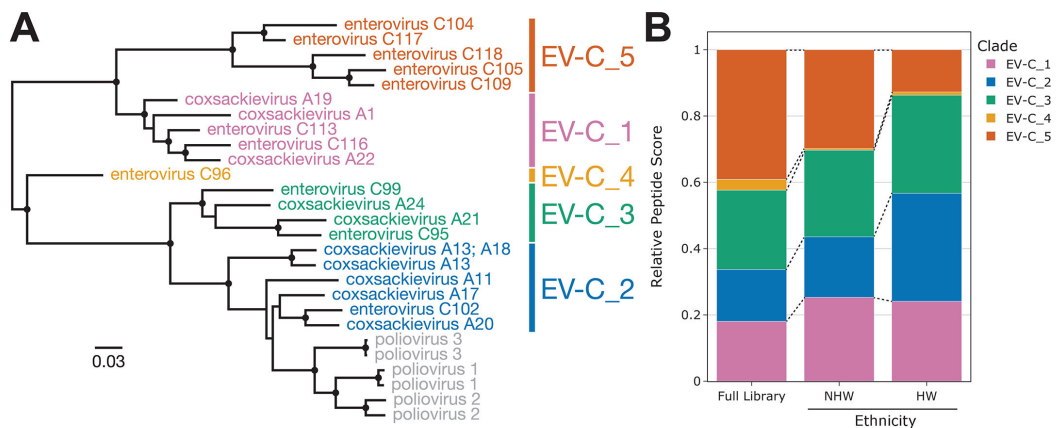


FIG 7 Antibody reactivity against clade-specific EV-C peptides. (A) Maximum-likelihood phylogenetic tree of all 27 ICTV listed EV-C isolates based on an amino acid alignment of the full polyprotein. Branch lengths indicate relative levels of amino acid divergence with the scale bar indicating the equivalent of 0.03 changes per site. Black circles indicate nodes with bootstrap support ≥ 80 . Six phylogenetic clades are identified by the different colors. GenBank accession numbers for each sequence can be found in Table S8. (B) Bar plot showing clade-specific relative peptide scores for all “Other EV-C” peptides present in the HV1 PepSeq library (Full Library, null distribution) compared to the subset of enriched peptides for all HW and NHW individuals predicted to be seropositive against “Other EV-C” (Fig. 6B). Each color represents one of the “Other EV-C” phylogenetic clades shown in panel A.

conformational epitopes; however, deviations from absolute seroprevalence are likely less important when focused on comparing between populations, as is done here, and, taken together, our results provide strong support for the use of our highly multiplexed serology approach for broadly characterizing virus infection histories.

Although we did not observe any statistically significant differences in seropositivity between males and females in our study, we did observe several nearly significant differences, all with a ~10% difference in seroprevalence between males and females (Fig. 5A). Notably, for two of these viruses (HIV-1 and HCV), our measured disparities are highly consistent with estimated disparities from medical records (PepSeq = 11%, 11%, and medical records = 10%, 10% for HIV-1 and HCV, respectively). In contrast, we observed moderately higher seroprevalence of Sapporo virus among females in our study. To our knowledge, this is the first time that such a disparity has been detected and the public health relevance is unknown. Sapporo virus infection is responsible for sporadic cases and outbreaks of diarrhea worldwide, primarily affecting children, older adults, and immunocompromised individuals (66).

We observed significant differences between HWs and NHWs for seven different viruses, and these differences were largely consistent across different Z score thresholds for peptide enrichment (Fig. 5B). For five of these viruses (CMV, HSV-1, HAV, SaV-A, and HAdV-D), we observed significantly higher seroprevalence within our HW subpopulation, while the other two (HHV-7 and PeV-A) showed higher seroprevalence in our NHW subpopulation, and several of these significant trends are consistent with published studies, lending additional credibility to our highly multiplexed approach. For example, we detected higher seroprevalence among HWs for both CMV and HSV-1 (both ~30% higher than among NHWs), consistent with the results of previous US population studies that used singleplex antibody assays (33, 44). We also observed significantly higher seroprevalence for HAV among HWs, which is consistent with published rates of seroprevalence in the US, as well as known differences in seroprevalence between the US and Mexico (35, 67–69).

Additionally, we observed several previously undocumented disparities in infection history between our HW and NHW populations, most of which involved viruses that are rarely targeted in serosurveys. One example is the recently described picornavirus SaV-A, which had ~20% higher seropositivity within our HW subpopulation. Little is known about the seroprevalence of SaV-A in the US, however, it has been documented in wastewater samples from multiple US states, and a study in Arizona found evidence of SaV-A in 15% of wastewater samples (70, 71). Another previously undocumented disparity involves HAdV-D, for which we observed ~25–30% higher seropositivity within our HW subpopulation (Fig. 5B). HAdV-D is a highly diverse species that has been shown to cause severe diseases such as epidemic keratoconjunctivitis in immunocompromised individuals. However, little is known about the general health impact of HAdV-D, and therefore, it is difficult to deduce the impact this virus might have on the well-being of this population (72). Finally, we estimated significantly higher seroprevalence (~15%–25%) in our NHW subpopulation for two understudied viruses: PeV-A and HHV-7. PeV-A is a widespread virus that is associated with respiratory and gastrointestinal symptoms (73, 74). However, certain strains of PeV-A are associated with more severe diseases, such as meningitis and sepsis-like illness in infants (74). HHV-7 is a highly prevalent virus, which is primarily contracted in early childhood, resulting in life-long latent infections that typically remain asymptomatic. However, HHV-7 has also been linked with febrile seizures and as a possible cause of encephalitis (75–77). More work is needed to understand the full impact that these undocumented disparities may have on human health within these populations.

Arizona is home to a large and diverse immigrant population and this has likely contributed to our observed differences in seropositivity among ethnicities. According to estimates from 2018, immigrants (i.e., foreign-born individuals) comprised 13% of the population in Arizona, 16% of the state's population were native-born Americans with at least one immigrant parent, and 55% of all immigrants in Arizona were from

Mexico, with the next most common countries of origin (Canada, India, Philippines) each accounting for only 4% of the immigrant population (78). Although we did not have access to the immigration status of the individuals included in this study, we were able to examine the payor source associated with each individual's medical visit, and we used a lack of insurance coverage (i.e., "self pay") as a proxy for immigration status. In other words, it was assumed that the uninsured population would contain a higher proportion of immigrants compared to the insured population, and consistent with this assumption, we observed a much higher proportion of uninsured individuals within our HW subpopulation compared to our NHW subpopulation (Fig. 1B). It is important to note, however, that the insurance coverage data is from a single time point and is not expected to fully represent the immigration status or socioeconomic background of the individuals in this study.

By comparing seroprevalence estimates between the uninsured and insured HW subpopulations, we were able to demonstrate that, for the majority of viruses that exhibited significant or near significant differences in seropositivity between ethnicities, the insured HW population represented an intermediate level of seropositivity between NHWs and uninsured HWs (Fig. 5C). This pattern is consistent with the hypothesis that a substantial portion of the differences in seroprevalence between ethnicities can be attributed to differences in seroprevalence between immigrant and non-immigrant populations. Antibody responses to viruses can be long-lived. Therefore, one possible explanation is that we are seeing evidence of differences in virus exposure rates, possibly during adolescence, for those who were raised in the US versus those who were raised outside of the US, potentially in less developed, resource-limited countries. Many of the viruses flagged by our analysis are commonly encountered during childhood, and several are known to be more prevalent outside of the US (CMV, HSV-1, and HAV) (79–81). However, other factors, such as differences in the dynamics of virus transmission in immigrant communities within the US could also be contributing to the observed patterns.

Virus transmission mechanisms may also play a role in determining which viruses are most likely to be associated with disparities. Among the seven viruses with a significant difference in seroprevalence between HWs and NHWs, three are primarily transmitted through intimate contact (CMV, HHV-7, and HSV-1) (82–84), three are primarily transmitted through the fecal-oral route (HAV, PeV-A, and SaV-A) (85–87), and one (HAdV-D) (88) has been associated with a variety of transmission routes, including close personal contact, the fecal-oral route, and respiratory droplets. Notably absent from this list are respiratory viruses that are primarily spread through airborne transmission. This absence of airborne transmitted viruses is likely due to the wider potential radius for spread from person to person and suggests that viruses that rely on close contact for successful transmission are more likely to be associated with population-level disparities.

Among our differentially seropositive species, two (HAV and EV-C) include viruses for which there are widely available vaccines, and for both of these, our highly multiplexed assay allowed us to assess the relative roles of vaccination and natural infection on the observed differences in seroprevalence. For HAV, we were able to differentiate antibody reactivity from vaccination and natural infection by leveraging the ability of PepSeq to simultaneously measure antibody reactivity across multiple protein targets. Specifically, we compared patterns of antibody reactivity between the structural and non-structural proteins of HAV (Fig. 6A). For non-structural proteins to be produced, active replication of viral particles must occur (89). However, all available vaccines in the US contain inactivated viruses, and therefore, vaccination will not elicit antibodies that target non-structural proteins. Our results show high levels of antibody reactivity against non-structural HAV proteins in both HW and NHW individuals. This suggests that the higher seroprevalence among HWs is not because of higher vaccination rates in this population.

There is also a commonly used vaccine that includes three viruses that belong to the EV-C species—poliovirus 1, 2, and 3—and both inactivated and live attenuated versions

were in use during the lifetimes of our participants. Therefore, to control for differences in antibody responses that may be driven by differences in rates of vaccination, we separated our EV-C peptides into two categories: those that share at least one 7-mer with any of the three polioviruses (i.e., those most similar to vaccine antigens) and those that do not. We saw no difference between our ethnicities in reactivity against the peptides most likely to be recognized by vaccine-induced antibodies, but a highly significant difference in reactivity against “Other EV-C” peptides (33% higher seropositivity among HWs; Fig. 6B). Further analysis showed that the observed antibody reactivity profiles are consistent with past exposures to a wide variety of EV-C viruses (Fig. S5), but that some phylogenetic clades may be contributing more than others to the observed disparity in EV-C infections between HWs and NHWs (Fig. 7B). Notably, while EV-C₅ peptides are overrepresented in our starting library, they are comparably underrepresented among our enriched peptides, especially in our HW subpopulation (Fig. 7B). Interestingly, EV-C₅ viruses (C104, C105, C109, C117, C118) have been predominantly isolated from nasal/throat swabs and nasopharyngeal aspirates, suggesting these viruses are likely transmitted via respiratory droplets (90, 91). In contrast, the other EV-C clades have been isolated almost universally from stool samples, indicating that they primarily infect the gastrointestinal tract and are likely to be spread through the fecal-oral transmission pathway (90, 91). This tissue-specific tropism aligns with our general hypothesis that the population-level virus infection disparities are more commonly driven by viruses that require close contact transmission routes.

Limitations

One limitation of the approach described herein is the inability to differentiate between a historical exposure and a recent infection. This study focused only on immunoglobulin G (IgG), which is typically generated within 2–3 weeks following viral infection, and in some cases, these responses can persist for a lifetime. Therefore, with the analysis presented here, it is generally not possible to determine the relative timing of responses. However, it is possible to augment the analysis utilizing longitudinal sampling and/or the capture of alternative isotypes, such as IgM, to obtain a more complete view of the timing of infections. Another factor to consider is the waning of antibody concentrations after an infection or vaccination event. The dynamics of virus-specific antibody populations after the infection is cleared are not well understood for most viruses. It is possible that the time from infection could play a role in the detection of a sample as seropositive. Another limitation is that we are unable to detect the full spectrum of antiviral antibodies using PepSeq, as some epitopes cannot be well-represented by peptides; for example, epitopes formed by tertiary and quaternary structures. For this study, however, neither of these factors is likely to have a large impact as (i) most viruses do elicit antibodies that are detectable in peptide-based assays (25), (ii) our estimates of seroprevalence are broadly consistent with expectations from published serology studies, and (iii) any limitations in sensitivity and/or resolution are expected to equally affect our different subpopulations.

Additionally, because of our sample size, relatively large differences in seroprevalence (>10%) were required for statistical significance. Future studies with larger sample sizes will be needed to detect smaller differences in seroprevalence that could still have a meaningful public health impact.

Conclusion

Overall, our highly multiplexed serology assay was successful in broadly characterizing antiviral antibody reactivities and allowed us to infer individual infection histories across the virome. The recapitulation of several known differences in seroprevalence between genders and ethnicities provides confidence in the quality of our analysis and highlights promising future applications for this type of approach. Additionally, our study revealed several previously undocumented disparities in virus seropositivity between HWs and NHWs in our study population. Future studies are needed to better understand

the clinical significance of these differences in infection rates and to develop medical and social interventions to minimize the impact of these disparities. Our results also demonstrate the potential for highly multiplexed serology to finely dissect the specificity and breadth of antibody responses, thus enabling an unprecedented view into an individual's history of infection. This pilot study demonstrates the potential for deploying this approach at a broader scale to conduct virome-wide seroprevalence studies and to determine community-level trends in infection rates.

MATERIALS AND METHODS

Study population and sample collection

In total, 400 serum samples were obtained from Valleywise Health in Phoenix, AZ. These samples were collected in late May and early June 2020. They were remnant samples initially collected as a part of the patient's standard of care and were collected from several different facilities and in a variety of contexts including outpatient encounters (55.5%), inpatient encounters (35%), and emergency department visits (9.5%). Researchers at Northern Arizona University (NAU) did not have access to any identifiable patient information.

To maximize statistical power to detect differences in seroprevalence, our cohort for this study was equally divided among four subpopulations: HW males ($n = 100$), HW females ($n = 100$), NHW males ($n = 100$), and NHW females ($n = 100$). To minimize the effect of age in detecting differences in seroprevalence, we selected only individuals within the age range of 30–60 years old. Self-reported ethnicity, gender, and age were the only characteristics considered for inclusion in this study, and we obtained aliquots for all eligible samples as they became available during our collection period until we met our target of 100 individuals per subpopulation. However, because Valleywise Health is a safety net hospital, we do not expect our study population to represent a random sampling of the population of Phoenix, AZ. Rather, it is likely to include a higher proportion of individuals with low income and from under-served populations. There is also some potential for bias associated with the use of remnant samples collected from patients actively receiving medical care (compared to a random sample of adults). However, given the wide variety of encounter types represented in our sample, we expect any impact to be minimal.

In total, 83% of the population served by Valleywise Health consists of racial and ethnic minorities, and at the Valleywise Health ambulatory clinics, 59% of patients are Hispanic. The majority of families served by Valleywise Health are at or below 150% of the federal poverty level, and approximately 55% of Valleywise Health patients are enrolled in a government health insurance program for low-income people or have insufficient private insurance or no insurance.

For each patient, we also obtained information regarding (i) HIV status, (ii) HCV status, and (iii) payor source. Both HIV-1 and HCV cause persistent infections in a high percentage of infected individuals (92, 93), and status was obtained by Valleywise Health staff by reviewing patient medical records. Payor sources can serve as a useful, though incomplete, indicator of socio-economic status. As the exact payor source is highly variable among individuals, we reduced the complexity of this categorical variable by assigning every individual to one of six general categories: (i) "Commercial," which included all commercial health plans; (ii) "Medicaid," which included both Arizona Health Care Cost Containment plans and out-of-state Medicaid; (iii) "Medicare"; (iv) "Dual-SNP," which included any dual special needs plans for individuals who qualify for both Medicaid and Medicare; (v) "Self Pay," for individuals without insurance; and (iv) "Other," which served as a final catch-all category that included funding through charitable organizations like the Ryan White HIV/AIDS Program and other government plans such as Tricare (Table S4).

Identification of published seroprevalence studies

A systematic search of PubMed for population-level studies of virus seroprevalence was conducted using the search terms (“United States;” “Virus;” and “Seroprevalence”). The resulting 267 hits were manually curated to select only papers analyzing population-level seroprevalence for viruses covered in our PepSeq assay in adults in the United States (Table S1).

PepSeq library design and assay

To broadly assess antiviral antibody reactivity, we utilized the PepSeq platform to perform highly multiplexed peptide-based serology (30). Specifically, we used our human virome version 1 (HV1) library described in reference 26. In brief, the HV1 library consists of 244,000 unique DNA-peptide conjugates (i.e., PepSeq probes). The variable peptide portion of each molecule is 30 amino acids long and the peptides were designed to broadly cover potential linear epitopes present in the proteins of viruses known to infect humans. Libraries of these PepSeq probes are created through a series of bulk, *in vitro* enzymatic reactions (30).

Each assay was conducted as described in Ladner et al. (26). Broadly, the PepSeq assay involves the incubation of serum with a diverse pool of PepSeq probes. IgG is then precipitated using magnetic protein G beads, non-binding PepSeq probes are washed away, and the relative abundance of each probe is quantified using PCR and high-throughput sequencing of the DNA portion of the molecules (30). Specifically, 5 μ L of a 1:10 dilution of serum in Superblock T20 (Thermo) was added to 0.1 pmol of the PepSeq library for a total volume of 10 μ L and was incubated at 20°C overnight. The binding reaction was incubated with pre-washed protein G-coated beads (Thermo) for 15 minutes, after which the beads were hand-washed 11 times with 1 \times PBST. After the final wash, beads were resuspended in 30 μ L of water and heated to 95°C for 5 minutes to elute the bound PepSeq probes. Elutions were amplified and indexed using barcoded DNA oligonucleotides (Table S5). Following PCR, a standard bead cleanup was performed and products were individually quantified (Quant-It, Thermo Fisher), pooled, re-quantified (KAPA Library Quantification Kit, Roche), and sequenced on a NextSeq instrument (Illumina). For this study, each sample was assayed in duplicate and ≥ 1 buffer-only negative controls were included on each 96-well assay plate. Potential batch effects were controlled through equal representation of each of our four focal subpopulations on each plate, as well as through the inclusion of negative controls from all plates in read count normalization and the generation of the peptide bins.

PepSeq analysis

We used PepSIRF v1.6.0 (94, 95) to analyze the high-throughput sequencing data. Demultiplexing and assignment of reads to peptides were done using the demux module of PepSIRF allowing up to one mismatch within each of the index sequences (12 and 8 nt, respectively) and up to three mismatches with the expected DNA tag (90 nt). Only samples that had two replicates with at least 2 \times sequencing depth per unique peptide ($\geq 488,000$ raw reads) were included in further analyses. Z scores were calculated using the zscore module of PepSIRF, which implements a method adapted from reference 96. This process involved the generation of peptide bins, each of which contained ≥ 300 peptides with similar expected abundances in our PepSeq library. The expected abundance for each peptide was estimated using buffer-only negative controls. In total, 12 independent buffer-only controls from seven different assay plates were used to generate the bins for this study. The raw read counts from each of these controls were first normalized to reads per million (RPM) using the column sum normalization method in the norm module of PepSIRF. This served to normalize for differences in total sequencing depth between samples. Bins were then generated using the bin PepSIRF module. RPM counts for each peptide were then further normalized by subtracting the average RPM count observed within our buffer-only controls. This second normalization step was used to control for any differences in initial relative abundance

among peptides contained within the same bin. Each Z score was calculated using peptides contained within the same bin and corresponds to the number of standard deviations away from the mean, with the mean and standard deviation calculated using the 95% highest density interval to exclude any enriched peptides. Only samples with a Z score Pearson's correlation ≥ 0.6 between replicates were included in the downstream analyses.

The enrich module of PepSIRF was used to determine which peptides had been enriched through our assay (i.e., were bound by serum IgG isotype antibodies). This module identifies peptides that meet or exceed minimum Z score thresholds, in both replicates for each sample. Z score thresholds were selected to minimize the number of false positive calls of peptide enrichment (determined through the analysis of negative controls that were not considered in the formation of bins), and multiple Z score thresholds were examined to determine the sensitivity of our results to changes in this threshold.

Estimating seropositivity from PepSeq

The lists of enriched peptides were converted into lists of putative species-level seropositivities using the deconv module of PepSIRF. The goal of this module is to predict the minimum list of viruses to which an individual has likely been infected while considering shared sequence diversity among different viruses. To accomplish this, the link module was first used to generate a linkage map that relates individual peptides to virus species. A link between a peptide and a virus indicates that enrichment of the peptide could be explained by exposure to the linked virus species, and these links were made whenever a peptide shared ≥ 1 amino acid 7-mer with a target protein sequence obtained from a particular virus species. Because of shared sequence diversity, a single peptide can be linked to multiple species, not just the species from which the peptide was designed. Furthermore, the strength of the link between peptides and viruses was quantified with scores that correspond to the number of shared 7-mers. Therefore, the maximum link score was 24 and the minimum link score was 1.

The deconv module of PepSIRF was then used with this linkage map to identify the most parsimonious set of virus species that can explain each set of enriched peptides. The results, therefore, can be interpreted as potential seropositivities for each sample. The deconv module accomplishes this through an iterative process. In each round, species-specific scores are generated by summing the species-level scores from each enriched peptide, and the species with the highest score is selected for inclusion in the output, as long as its score meets or exceeds a minimum score threshold. Additionally, to account for related species with similar scores, we allowed for ties between species if the lower scoring species had a score $\geq 80\%$ of the higher scoring species (`--score_tie_threshold 0.8`) and if $\geq 70\%$ of the enriched peptides contributing to these scores were identical between species (`--score_overlap_threshold 0.7`). In our analysis, we considered an individual to be seropositive for all tied species. Ties accounted for less than 2.3% of seropositivity calls.

Initially, we tested a range of seropositivity score thresholds (`--threshold 20, 40, 60, 200, 600`; thresholds were fixed across all virus species), in combination with a Z score threshold of 15, to determine which would work best across the diversity of viruses covered by our assay. For each threshold, we compared our PepSeq-based estimates of seropositivity to those from published serosurveillance studies conducted on adults in the US (see above). Using the optimal seropositivity score threshold (i.e., best congruence with published seroprevalence estimates), we calculated deviations between PepSeq-based estimates of seroprevalence and average published estimates of seroprevalence for each virus. We then compared these deviations to the total possible seropositivity score for each virus, which provides an indication of the relative level of representation in our library and was calculated by summing the link scores across all 244,000 peptides for each virus species in the HV1 linkage map.

To normalize representation bias in our HV1 library, we reran *deconv* using virus-specific seropositivity score thresholds based on each virus' total possible score. To calculate the species-specific seropositivity thresholds, we empirically determined optimal score thresholds for three viruses by comparing our PepSeq results to the results from singleplex ELISAs (see below). We then fit a curve relating the optimal score threshold to the total possible seropositivity score using a third-order polynomial and enforcing a plateau at a minimum score threshold of 20 (Fig. S3B). The equation from this curve was then used to select representation-normalized score thresholds for all 390 species covered by HV1 (Table S6). A separate *deconv* analysis was run for each Z score threshold (Table S7).

ELISA assays

For a randomly selected subset of our collected serum samples ($n = 78\text{--}87$) and three focal viruses, we used ELISA assays to measure seropositivity independent of our PepSeq assay. For this benchmarking analysis, we focused on three viruses with commercially available ELISA assays that are expected (from published serosurveys) to exhibit intermediate levels of seropositivity in the US adult population: CMV (expected seroprevalence = ~64%), HSV-1 (~60%–67%), and HSV-2 (~16%–25%). Human IgG ELISA kits were purchased from Abcam (CMV: ab108724, HSV-1: ab108737, HSV-2: ab108739). All assays were conducted by following the manufacturer's recommended protocol. Each sample and control (positive, negative, cutoff, and blank on each plate) was assayed in duplicate and the absorbance was averaged across replicates (Table S2). A sample was considered negative if its average absorbance was less than 90% of the cutoff value and positive if its absorbance was >110% of the cutoff value; otherwise, the test was considered inconclusive. Inconclusive samples were dropped from further analysis.

Receiver operating characteristic curves

To further assess the impact of the seropositivity score threshold on the sensitivity and specificity of our PepSeq assay, we compared our PepSeq-based seropositivity calls to the results of the traditional singleplex ELISA assays and documented HIV-1/HCV infection status using ROC curves. The ELISA or known infection data were utilized as truth and *deconv* was run with seropositivity score thresholds ranging from 0 to 14,000 for each of the four Z score thresholds (10, 15, 20, and 25). Sensitivity and specificity were calculated for each combination of score thresholds, and the optimal seropositivity score threshold for each virus was selected by maximizing the sum of sensitivity and specificity (averaged across Z score thresholds).

Identification of disparities

To identify significant differences in estimated seroprevalence between ethnicities and/or genders, we utilized a GLM implemented in Python using *statsmodels* v3.8.8 (97). Specifically, for each viral species, we fit a binomial GLM with a single dependent variable (seropositivity; 0 or 1 for each individual) and three independent variables (ethnicity [categorical], gender [categorical], and age [continuous]). We utilized an alpha of 0.05 for determining significance, along with a Bonferroni correction for multiple tests (i.e., number of viruses). To reduce the total number of tests, we only examined viruses with estimated population-level seropositivities between 5% and 95%. These thresholds were chosen based on power simulations, which indicated that, with our sample size, it would be unlikely to detect differences in seropositivity <10% as statistically significant.

Subspecies analysis of reactivity profiles

To assign enriched HAV peptides to individual proteins, we first aligned all 360 HAV amino acid sequences from which HV1 peptides were designed using *mafft* v7.490 (98) with the *G-INS-i method. Annotations from one of these sequences (Uniprot:Q9DWR1) were then translated into alignment coordinates for visualization (Fig. 6A) and for

assigning peptides to proteins. A peptide was considered structural if ≥ 25 peptide amino acids (83%) were assigned to HAV proteins VP1-4 and/or 2A (blue in Fig. 6A). A peptide was considered non-structural if ≥ 25 peptide amino acids were assigned to HAV proteins 2BC and/or 3ABCD (green in Fig. 6A).

To determine if the disparity in EV-C seropositivity was driven by differences in vaccination rate, we created a new linkage map where all peptides containing at least one 7-mer from any of the 1,276 poliovirus sequences (Table S3) used in the creation of the HV1 library were assigned to a new taxonomic category called "poliovirus." All EV-C peptides that did not share a 7-mer with poliovirus sequences were assigned to an "Other EV-C" category. Next, we reran the deconv module using this modified linkage map to estimate seropositivity for these two categories separately. With this analysis, it is possible for an individual sample to be found seropositive for (i) just one of these categories, (ii) both categories (if there are enriched peptides from each category), or (iii) neither of these categories (even for samples previously found to be seropositive for EV-C, if the enriched peptides are split between the new categories). This is a conservative approach for assessing the impact of poliovirus vaccination/infection because it assumes that any antibody recognizing a peptide that shares at least one amino acid 7-mer with poliovirus was generated in response to poliovirus infection/vaccination. In reality, these responses could have been stimulated by many different EV-C strains.

To examine the contribution of different strains of EV-C to our observed "Other EV-C" seropositivities, we assigned each "Other EV-C" peptide to a single subspecies group based on shared amino acid 7-mers, with each peptide assigned a score (between 1 and 24) equivalent to the number of contained 7-mers that were unique to the assigned group. To improve the sensitivity of our analysis (i.e., the number of informative peptides), we focused on clades of related EV-C isolates. As of 8 December 2023, the ICTV website for the Enterovirus genus (<https://ictv.global/report/chapter/picornaviridae/picornaviridae/enterovirus>) listed NCBI GenBank accession numbers for 27 EV-C isolates, including six poliovirus isolates (Table S8). We downloaded polyprotein amino acid sequences for each of these and aligned them using mafft v7.490 (98) with default settings. We generated a maximum-likelihood phylogeny from this alignment using raxml-ng v0.5.1b (99) with the LG+FC+I+G8m model, which was selected as optimal using ModelGenerator v0.85 (AIC1) (100). Based on this phylogeny, we divided the EV-C subspecies into six phylogenetic clades, including five composed of non-poliovirus EV-C (EV-C_1-5) (Fig. 7). For each enriched "Other EV-C" peptide, we determined the number of amino acid 7-mers shared with the ICTV reference sequences from each EV-C clade. We assigned the peptide to the clade with the highest score and normalized the associated score for this peptide by subtracting the next highest clade-specific score (no assignment was made if two clades had identical 7-mer scores). To ensure our analysis focused on samples with substantial EV-C clade-specific signal, only samples that had a total clade-specific score (across all five clades) greater than 20 were included in the results summary. To calculate relative reactivity scores against the five non-poliovirus clades ("Relative Peptide Score" in Fig. 7B; Fig. S5), each clade-specific sum of enriched peptide scores was divided by the total sum of scores across all clade-assigned peptides for that sample. For the ethnicity-level composite scores shown in Fig. 7B ("HW" and "NHW"), we summed enriched peptide scores across all HW or NHW individuals, respectively, who were seropositive for our "Other EV-C" category. For these composites, individual peptide scores were counted once for every sample that exhibited enrichment (i.e., a single peptide could be counted multiple times if that peptide was recognized by antibodies in multiple samples).

ACKNOWLEDGMENTS

We would like to acknowledge Sarah Namdarian for helping to collect the clinical samples used in this study.

This work was supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health (NIH) under award number U54MD012388,

the National Institute of Allergy and Infectious Diseases of the NIH under award number U24AI152172, and the State of Arizona Technology and Research Initiative Fund (TRIF; administered by the Arizona Board of Regents, through Northern Arizona University). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

AUTHOR AFFILIATIONS

¹The Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, Arizona, USA

²Valleywise Health, Phoenix, Arizona, USA

³University of Arizona, College of Medicine, Phoenix, Arizona, USA

⁴Creighton University, School of Medicine, Phoenix, Arizona, USA

⁵The Translational Genomics Research Institute (TGen), Flagstaff, Arizona, USA

AUTHOR ORCID*s*

Evan A. Elko  <http://orcid.org/0000-0001-6472-3169>

Jason T. Ladner  <http://orcid.org/0000-0001-7751-4366>

FUNDING

Funder	Grant(s)	Author(s)
HHS National Institutes of Health (NIH)	U54MD012388	John A. Altin Jason T. Ladner
HHS National Institutes of Health (NIH)	U24AI152172	John A. Altin Jason T. Ladner

AUTHOR CONTRIBUTIONS

Alejandra Piña, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing | Evan A. Elko, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing | Rachel Caballero, Data curation, Resources | Morgan Metrailler, Formal analysis, Investigation, Validation | Mary Mulrow, Data curation, Resources | Dan Quan, Supervision, Writing – review and editing | Lora Nordstrom, Data curation, Investigation, Project administration, Resources, Supervision, Writing – review and editing | John A. Altin, Funding acquisition, Methodology, Writing – review and editing | Jason T. Ladner, Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Software, Supervision, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The raw peptide counts, linkage maps, and data related to the figures from this study have been deposited in the Open Science Framework (<https://osf.io/gvpzu/>), DOI: [10.17605/OSF.IO/GVPZU](https://doi.org/10.17605/OSF.IO/GVPZU). All custom code is available via GitHub (<https://github.com/LadnerLab>). Any additional information required to reanalyze the data reported in this paper is available upon request.

ETHICS APPROVAL

This study was reviewed and approved by the Valleywise Health and NAU Institutional Review Boards (approval number 1545420).

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental Figures (mSphere00127-24-s0001.pdf). Figures S1 to S6.
Supplemental Tables (mSphere00127-24-s0002.xlsx). Tables S1 to S8.

REFERENCES

1. Tyler KL. 2018. Acute viral encephalitis. *N Engl J Med* 379:557–566. <https://doi.org/10.1056/NEJMra1708714>
2. Naniche D, Oldstone MB. 2000. Generalized immunosuppression: how viruses undermine the immune response. *Cell Mol Life Sci* 57:1399–1407. <https://doi.org/10.1007/PL00000625>
3. Pereira L. 2018. Congenital viral infection: traversing the uterine-placental interface. *Annu Rev Virol* 5:273–299. <https://doi.org/10.1146/annurev-virology-092917-043236>
4. Lu S, Huang X, Liu R, Lan Y, Lei Y, Zeng F, Tang X, He H. 2022. Comparison of COVID-19 induced respiratory failure and typical ards: similarities and differences. *Front Med (Lausanne)* 9:829771. <https://doi.org/10.3389/fmed.2022.829771>
5. Shahrizaila N, Lehmann HC, Kuwabara S. 2021. Guillain-Barré syndrome. *Lancet* 397:1214–1228. [https://doi.org/10.1016/S0140-6736\(21\)00517-1](https://doi.org/10.1016/S0140-6736(21)00517-1)
6. Feldstein LR, Rose EB, Horwitz SM, Collins JP, Newhams MM, Son MBF, Newburger JW, Kleinman LC, Heidemann SM, Martin AA, et al. 2020. Multisystem inflammatory syndrome in U.S. children and adolescents. *N Engl J Med* 383:334–346. <https://doi.org/10.1056/NEJMoa2021680>
7. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, Pujol JC, Klaser K, Antonelli M, Canas LS, et al. 2021. Attributes and predictors of long COVID. *Nat Med* 27:626–631. <https://doi.org/10.1038/s41591-021-01292-y>
8. Tobin NH, Campbell AJP, Zerr DM, Melvin AJ. 2011. Chapter 95, Life-threatening viral diseases and their treatment, p 1324–1335. In Fuhrman BP, Zimmerman JJ (ed), *Pediatric critical care*, 4th ed. Mosby, Saint Louis.
9. Filippi CM, von Herrath MG. 2008. Viral trigger for type 1 diabetes: pros and cons. *Diabetes* 57:2863–2871. <https://doi.org/10.2337/db07-1023>
10. Brown JJ, Jabri B, Dermody TS. 2018. A viral trigger for celiac disease. *PLoS Pathog* 14:e1007181. <https://doi.org/10.1371/journal.ppat.1007181>
11. Mitra AK, Clarke K. 2010. Viral obesity: fact or fiction? *Obes Rev* 11:289–296. <https://doi.org/10.1111/j.1467-789X.2009.00677.x>
12. Bar-Or A, Pender MP, Khanna R, Steinman L, Hartung H-P, Maniar T, Croze E, Aftab BT, Giovannoni G, Joshi MA. 2020. Epstein-Barr virus in multiple sclerosis: theory and emerging immunotherapies. *Trends Mol Med* 26:296–310. <https://doi.org/10.1016/j.molmed.2019.11.003>
13. Bjernevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, Elledge SJ, Niebuhr DW, Scher AI, Munger KL, Ascherio A. 2022. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* 375:296–301. <https://doi.org/10.1126/science.abj8222>
14. Sarid R, Gao S-J. 2011. Viruses and human cancer: from detection to causality. *Cancer Lett* 305:218–227. <https://doi.org/10.1016/j.canlet.2010.09.011>
15. Devanand DP. 2018. Viral hypothesis and antiviral treatment in Alzheimer's disease. *Curr Neurol Neurosci Rep* 18:55. <https://doi.org/10.1007/s11910-018-0863-1>
16. Liu G, Markowitz LE, Hariri S, Panicker G, Unger ER. 2016. Seroprevalence of 9 human papillomavirus types in the United States, 2005–2006. *J Infect Dis* 213:191–198. <https://doi.org/10.1093/infdis/jiv403>
17. Xu F, Sternberg MR, Kottiri BJ, McQuillan GM, Lee FK, Nahmias AJ, Berman SM, Markowitz LE. 2006. Trends in herpes simplex virus type 1 and type 2 seroprevalence in the United States. *JAMA* 296:964–973. <https://doi.org/10.1001/jama.296.8.964>
18. CDC. 2021. HIV and gay and bisexual men. Available from: <https://www.cdc.gov/hiv/group/msm/index.html>. Retrieved 15 Aug 2022.
19. Purcell DW, Johnson CH, Lansky A, Prejean J, Stein R, Denning P, Gau Z, Weinstock H, Su J, Crepaz N. 2012. Estimating the population size of men who have sex with men in the United States to obtain HIV and syphilis rates. *Open AIDS J* 6:98–107. <https://doi.org/10.2174/1874613601206010098>
20. Erhart LM, Ernst KC. 2012. The changing epidemiology of hepatitis A in Arizona following intensive immunization programs (1988–2007). *Vaccine* 30:6103–6110. <https://doi.org/10.1016/j.vaccine.2012.07.029>
21. World Health Organization. 2009. Dengue: guidelines for diagnosis, treatment, prevention and control. World Health Organization
22. Statistics and maps - 2020. 2022. Available from: <https://www.cdc.gov/dengue/statistics-maps/2020.html>. Retrieved 23 Aug 2022.
23. Hans R, Marwaha N. 2014. Nucleic acid testing-benefits and constraints. *Asian J Transfus Sci* 8:2–3. <https://doi.org/10.4103/0973-6247.126679>
24. Lequin RM. 2005. Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clin Chem* 51:2415–2418. <https://doi.org/10.1373/clinchem.2005.051532>
25. Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung'u T, Ruxrungtham K, Sanchez J, Brander C, Chung RT, O'Connor KC, Walker B, Larman HB, Elledge SJ. 2015. Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. *Science* 348:aaa0698. <https://doi.org/10.1126/science.aaa0698>
26. Ladner JT, Henson SN, Boyle AS, Engelbrekton AL, Fink ZW, Rahee F, D'ambrozio J, Schaecher KE, Stone M, Dong W, Dadwal S, Yu J, Caligiuri MA, Cieplak P, Björås M, Fenstad MH, Nordbø SA, Kainov DE, Muranaka N, Chee MS, Shiryayev SA, Altin JA. 2021. Epitope-resolved profiling of the SARS-CoV-2 antibody response identifies cross-reactivity with endemic human coronaviruses. *Cell Rep Med* 2:100189. <https://doi.org/10.1016/j.xcrm.2020.100189>
27. Elko EA, Nelson GA, Mead HL, Kelley EJ, Carvalho ST, Sarbo NG, Harms CE, Le Verche V, Cardoso AA, Ely JL, Boyle AS, Piña A, Henson SN, Rahee F, Keim PS, Celona KR, Yi J, Settles EW, Bota DA, Yu GC, Morris SR, Zaia JA, Ladner JT, Altin JA. 2022. COVID-19 vaccination elicits an evolving, cross-reactive antibody response to epitopes conserved with endemic coronavirus spike proteins. *Cell Rep* 40:111022. <https://doi.org/10.1016/j.celrep.2022.111022>
28. Kelley EJ, Henson SN, Rahee F, Boyle AS, Engelbrekton AL, Nelson GA, Mead HL, Anderson NL, Razavi M, Yip R, Ladner JT, Scriba TJ, Altin JA. 2023. Virome-wide detection of natural infection events and the associated antibody dynamics using longitudinal highly-multiplexed serology. *Nat Commun* 14:1783. <https://doi.org/10.1038/s41467-023-37378-z>
29. Elko EA, Mead HL, Nelson GA, Zaia JA, Ladner JT, Altin JA. 2024. Recurrent SARS-CoV-2 mutations at Spike D796 evade antibodies from pre-Omicron convalescent and vaccinated subjects. *Microbiol Spectr* 12:e0329123. <https://doi.org/10.1128/spectrum.03291-23>
30. Henson SN, Elko EA, Swiderski PM, Liang Y, Engelbrekton AL, Piña A, Boyle AS, Fink Z, Facista SJ, Martinez V, Rahee F, Brown A, Kelley EJ, Nelson GA, Raspet I, Mead HL, Altin JA, Ladner JT. 2023. PepSeq: a fully *in vitro* platform for highly multiplexed serology using customizable DNA-barcoded peptide libraries. *Nat Protoc* 18:396–423. <https://doi.org/10.1038/s41596-022-00766-8>
31. Mohan D, Wansley DL, Sie BM, Noon MS, Baer AN, Laserson U, Larman HB. 2018. Phip-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nat Protoc* 13:1958–1978. <https://doi.org/10.1038/s41596-018-0025-6>
32. Patricia O'Campo JB. 2004. Eliminating health disparities: measurement and data needs. National Academies Press.
33. Staras SAS, Dollard SC, Radford KW, Flanders WD, Pass RF, Cannon MJ. 2006. Seroprevalence of cytomegalovirus infection in the United States, 1988–1994. *Clin Infect Dis* 43:1143–1151. <https://doi.org/10.1086/508173>
34. Balfour HH, Sifakis F, Sliman JA, Knight JA, Schmeling DO, Thomas W. 2013. Age-specific prevalence of Epstein-Barr virus infection among individuals aged 6–19 years in the United States and factors affecting its acquisition. *J Infect Dis* 208:1286–1293. <https://doi.org/10.1093/infdis/jit321>

35. Klevens RM, Kruszon-Moran D, Wasley A, Gallagher K, McQuillan GM, Kuhnhert W, Teshale EH, Drobeniuc J, Bell BP. 2011. Seroprevalence of hepatitis A virus antibodies in the U.S.: results from the national health and nutrition examination survey. *Public Health Rep* 126:522–532. <https://doi.org/10.1177/003335491112600408>
36. Bell BP, Kruszon-Moran D, Shapiro CN, Lambert SB, McQuillan GM, Margolis HS. 2005. Hepatitis A virus infection in the United States: serologic results from the third national health and nutrition examination survey. *Vaccine* 23:5798–5806. <https://doi.org/10.1016/j.vaccine.2005.03.060>
37. McQuillan GM, Kruszon-Moran D, Kottiri BJ, Curtin LR, Lucas JW, Kington RS. 2004. Racial and ethnic differences in the seroprevalence of 6 infectious diseases in the United States: data from NHANES III, 1988–1994. *Am J Public Health* 94:1952–1958. <https://doi.org/10.2105/ajph.94.11.1952>
38. Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, Kaslow RA, Margolis HS. 1999. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med* 341:556–562. <https://doi.org/10.1056/NEJM199908193410802>
39. Okuno T, Takahashi K, Balachandra K, Shiraki K, Yamanishi K, Takahashi M, Baba K. 1989. Seroepidemiology of human herpesvirus 6 infection in normal children and adults. *J Clin Microbiol* 27:651–653. <https://doi.org/10.1128/jcm.27.4.651-653.1989>
40. Engels EA, Atkinson JO, Graubard BI, McQuillan GM, Gamache C, Mbisa G, Cohn S, Whitby D, Goedert JJ. 2007. Risk factors for human herpesvirus 8 infection among adults in the United States and evidence for sexual transmission. *J Infect Dis* 196:199–207. <https://doi.org/10.1086/518791>
41. Louis Mes, Rauch KJ, Petersen LR, Anderson JE, Schable CA, Dondero TJ. 1990. Seroprevalence rates of human immunodeficiency virus infection at sentinel hospitals in the United States. The sentinel hospital surveillance group. *N Engl J Med* 323:213–218. <https://doi.org/10.1056/NEJM199007263230401>
42. Markowitz LE, Sternberg M, Dunne EF, McQuillan G, Unger ER. 2009. Seroprevalence of human papillomavirus types 6, 11, 16, and 18 in the United States: national health and nutrition examination survey 2003–2004. *J Infect Dis* 200:1059–1067. <https://doi.org/10.1086/604729>
43. Schillinger JA, Xu F, Sternberg MR, Armstrong GL, Lee FK, Nahmias AJ, McQuillan GM, Louis ME, Markowitz LE. 2004. National seroprevalence and trends in herpes simplex virus type 1 in the United States, 1976–1994. *Sex Transm Dis* 31:753–760. <https://doi.org/10.1097/01.oq.0000145852.43262.c3>
44. McQuillan G, Kruszon-Moran D, Flagg EW, Paulose-Ram R. 2018. Prevalence of herpes simplex virus type 1 and type 2 in persons aged 14–49: United States, 2015–2016. *NCHS Data Brief*:1–8.
45. Fleming DT, McQuillan GM, Johnson RE, Nahmias AJ, Aral SO, Lee FK, St Louis ME. 1997. Herpes simplex virus type 2 in the United States, 1976 to 1994. *N Engl J Med* 337:1105–1111. <https://doi.org/10.1056/NEJM199710163371601>
46. Centers for Disease Control and Prevention (CDC). 2010. Seroprevalence of herpes simplex virus type 2 among persons aged 14–49 years—United States, 2005–2008. *MMWR Morb Mortal Wkly Rep* 59:456–459.
47. McQuillan GM, Coleman PJ, Kruszon-Moran D, Moyer LA, Lambert SB, Margolis HS. 1999. Prevalence of hepatitis B virus infection in the United States: the national health and nutrition examination surveys, 1976 through 1994. *Am J Public Health* 89:14–18. <https://doi.org/10.2105/ajph.89.1.14>
48. Teshale EH, Denniston MM, Drobeniuc J, Kamili S, Teo C-G, Holmberg SD. 2015. Decline in hepatitis E virus antibody prevalence in the United States from 1988–1994 to 2009–2010. *J Infect Dis* 211:366–373. <https://doi.org/10.1093/infdis/jiu466>
49. Kuniholm MH, Purcell RH, McQuillan GM, Engle RE, Wasley A, Nelson KE. 2009. Epidemiology of hepatitis E virus in the United States: results from the third national health and nutrition examination survey, 1988–1994. *J Infect Dis* 200:48–56. <https://doi.org/10.1086/599319>
50. Lebo EJ, Kruszon-Moran DM, Marin M, Bellini WJ, Schmid S, Bialek SR, Wallace GS, McLean HQ. 2015. Seroprevalence of measles, mumps, rubella and varicella antibodies in the United States population, 2009–2010. *Open Forum Infect Dis* 2:fv006. <https://doi.org/10.1093/ofid/ofv006>
51. Hutchins SS, Bellini WJ, Coronado V, Jiles R, Wooten K, Deladisma A. 2004. Population immunity to measles in the United States, 1999. *J Infect Dis* 189 Suppl 1:S91–S97. <https://doi.org/10.1086/377713>
52. Kirby AE, Kienast Y, Zhu W, Barton J, Anderson E, Sizemore M, Vinje J, Moe CL. 2020. Norovirus seroprevalence among adults in the United States: analysis of NHANES serum specimens from 1999–2000 and 2003–2004. *Viruses* 12:179. <https://doi.org/10.3390/v12020179>
53. Pisano MB, Campbell C, Anugwom C, Ré VE, Debes JD. 2022. Hepatitis E virus infection in the United States: seroprevalence, risk factors and the influence of immunological assays. *PLoS One* 17:e0272809. <https://doi.org/10.1371/journal.pone.0272809>
54. Cahill ME, Yao Y, Nock D, Armstrong PM, Andreadis TG, Diuk-Wasser MA, Montgomery RR. 2017. West Nile virus seroprevalence, Connecticut, USA, 2000–2014. *Emerg Infect Dis* 23:708–710. <https://doi.org/10.3201/eid2304.161669>
55. Schweitzer BK, Kramer WL, Sambol AR, Meza JL, Hinrichs SH, Iwen PC. 2006. Geographic factors contributing to a high seroprevalence of West Nile virus-specific antibodies in humans following an epidemic. *Clin Vaccine Immunol* 13:314–318. <https://doi.org/10.1128/CI.13.3.314-318.2006>
56. Ye C, Luo J, Wang X, Xi J, Pan Y, Chen J, Yang X, Li G, Sun Q, Yang J. 2017. Development of A peptide ELISA to discriminate vaccine-induced immunity from natural infection of hepatitis A virus in A phase IV study. *Eur J Clin Microbiol Infect Dis* 36:2165–2170. <https://doi.org/10.1007/s10096-017-3040-6>
57. Alfaro-Murillo JA, Ávila-Agüero ML, Fitzpatrick MC, Crystal CJ, Falleiros-Arlant L-H, Galvani AP. 2020. The case for replacing live oral polio vaccine with inactivated vaccine in the Americas. *Lancet* 395:1163–1166. [https://doi.org/10.1016/S0140-6736\(20\)30213-0](https://doi.org/10.1016/S0140-6736(20)30213-0)
58. Kang X, Li Y, Fan L, Lin F, Wei J, Zhu X, Hu Y, Li J, Chang G, Zhu Q, Liu H, Yang Y. 2012. Development of an ELISA-array for simultaneous detection of five encephalitis viruses. *Viol J* 9:56. <https://doi.org/10.1186/1743-422X-9-56>
59. Fukushi S. 2020. Competitive ELISA for the detection of serum antibodies specific for middle east respiratory syndrome coronavirus (MERS-CoV). *Methods Mol Biol* 2203:55–65. https://doi.org/10.1007/978-1-0716-0900-2_4
60. Cangin C, Focht B, Harris R, Strunk JA. 2019. Hepatitis E seroprevalence in the United States: results for immunoglobulins IGG and IGM. *J Med Virol* 91:124–131. <https://doi.org/10.1002/jmv.25299>
61. Boncristiani HF, Criado MF, Arruda E. 2009. Respiratory viruses, p 500–518. In *Encyclopedia of microbiology*. Elsevier.
62. Dunn SR, Ryder AB, Tollefson SJ, Xu M, Saville BR, Williams JV. 2013. Seroepidemiologies of human metapneumovirus and respiratory syncytial virus in young children, determined with a new recombinant fusion protein enzyme-linked immunosorbent assay. *Clin Vaccine Immunol* 20:1654–1656. <https://doi.org/10.1128/CI.00750-12>
63. Parrott RH, Vargosko AJ, BellJA, ChanockRM. 1962. Acute respiratory diseases of viral etiology. III. parainfluenza. Myxoviruses. *Am J Public Health Nations Health* 52:907–917. <https://doi.org/10.2105/ajph.52.6.907>
64. DeGroot NP, Haynes AK, Taylor C, Killerby ME, Dahl RM, Mustaqim D, Gerber SI, Watson JT. 2020. Human parainfluenza virus circulation, United States, 2011–2019. *J Clin Virol* 124:104261. <https://doi.org/10.1016/j.jcv.2020.104261>
65. Perrone O, Meissner HC. 2020. The importance of MMR immunization in the United States. *Pediatrics* 146:e20200251. <https://doi.org/10.1542/peds.2020-0251>
66. Becker-Dreps S, González F, Bucardo F. 2020. Sapovirus: an emerging cause of childhood diarrhea. *Curr Opin Infect Dis* 33:388–397. <https://doi.org/10.1097/QCO.0000000000000671>
67. Rubicz R, Leach CT, Kraig E, Dhurandhar NV, Grubbs B, Blangero J, Yolken R, Göring HH. 2011. Seroprevalence of 13 common pathogens in a rapidly growing U.S. minority population: Mexican Americans from San Antonio, TX. *BMC Res Notes* 4:433. <https://doi.org/10.1186/1756-0500-4-433>
68. Ernst KC, Erhart LM. 2014. The role of ethnicity and travel on hepatitis A vaccination coverage and disease incidence in Arizona at the United States-Mexico Border. *Hum Vaccin Immunother* 10:1396–1403. <https://doi.org/10.4161/hv.28140>

69. Lazcano-Ponce E, Conde-Gonzalez C, Rojas R, DeAntonio R, Romano-Mazzotti L, Cervantes Y, Ortega-Barria E. 2013. Seroprevalence of hepatitis A virus in a cross-sectional study in Mexico: implications for hepatitis A vaccination. *Hum Vaccin Immunother* 9:375–381. <https://doi.org/10.4161/hv.22774>
70. Li L, Victoria J, Kapoor A, Blinkova O, Wang C, Babrzadeh F, Mason CJ, Pandey P, Triki H, Bahri O, Oderinde BS, Baba MM, Bukbuk DN, Besser JM, Bartkus JM, Delwart EL. 2009. A novel picornavirus associated with gastroenteritis. *J Virol* 83:12002–12006. <https://doi.org/10.1128/JVI.01241-09>
71. Kitajima M, Iker BC, Rachmadi AT, Haramoto E, Gerba CP. 2014. Quantification and genetic analysis of salivirus/klasevirus in wastewater in Arizona, USA. *Food Environ Virol* 6:213–216. <https://doi.org/10.1007/s12560-014-9148-2>
72. Pauly M, Hoppe E, Mugisha L, Petrzekova K, Akoua-Koffi C, Couacy-Hymann E, Anoh AE, Mossou A, Schubert G, Wiersma L, Pascale S, Muyembe J-J, Karhemere S, Weiss S, Leendertz SA, Calvignac-Spencer S, Leendertz FH, Ehlers B. 2014. High prevalence and diversity of species D adenoviruses (HAdV-D) in human populations of four Sub-Saharan countries. *Virol J* 11:25. <https://doi.org/10.1186/1743-422X-11-25>
73. Karelehto E, Brouwer L, Benschop K, Kok J, Basile K, McMullan B, Rawlinson W, Druce J, Nicholson S, Selvarangan R, Harrison C, Lankachandra K, van Eijk H, Koen G, de Jong M, Pajkrt D, Wolthers KC. 2019. Seroepidemiology of parechovirus A3 neutralizing antibodies, Australia, the Netherlands, and United States. *Emerg Infect Dis* 25:148–152. <https://doi.org/10.3201/eid2501.180352>
74. Sridhar A, Karelehto E, Brouwer L, Pajkrt D, Wolthers KC. 2019. Parechovirus A pathogenesis and the enigma of genotype A-3. *Viruses* 11:1062. <https://doi.org/10.3390/v11111062>
75. Oliver S, James SH. 2017. Herpesviruses, p 621–627. In *International encyclopedia of public health*. Elsevier.
76. Epstein LG, Shinnar S, Hesdorffer DC, Nordli DR, Hamidullah A, Benn EKT, Pellock JM, Frank LM, Lewis DV, Moshe SL, Shinnar RC, Sun S, FEBSTAT study team. 2012. Human herpesvirus 6 and 7 in febrile status epilepticus: the FEBSTAT study. *Epilepsia* 53:1481–1488. <https://doi.org/10.1111/j.1528-1167.2012.03542.x>
77. Li Y, Qu T, Li D, Jing J, Deng Q, Wan X. 2022. Human herpesvirus 7 encephalitis in an immunocompetent adult and a literature review. *Virol J* 19:200. <https://doi.org/10.1186/s12985-022-01925-9>
78. Immigrants in Arizona. 2020. American Immigration Council.
79. Wald A, Corey L. 2007 Persistence in the population: epidemiology, transmission. In Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (ed), *Human herpesviruses: biology, therapy, and immunoprophylaxis*. Cambridge University Press, Cambridge.
80. Cao G, Jing W, Liu J, Liu M. 2021. The global trends and regional differences in incidence and mortality of hepatitis A from 1990 to 2019 and implications for its prevention. *Hepatol Int* 15:1068–1082. <https://doi.org/10.1007/s12072-021-10232-4>
81. Zuhair M, Smit GSA, Wallis G, Jabbar F, Smith C, Devleeschauwer B, Griffiths P. 2019. Estimation of the worldwide seroprevalence of cytomegalovirus: a systematic review and meta-analysis. *Rev Med Virol* 29:e2034. <https://doi.org/10.1002/rmv.2034>
82. Dioverti MV, Razonable RR. 2016. Cytomegalovirus. *Microbiol Spectr* 4. <https://doi.org/10.1128/microbiolspec.DMIH2-0022-2015>
83. Verbeek R, Vandekerckhove L, Van Cleemput J. 2024. Update on human herpesvirus 7 pathogenesis and clinical aspects as a roadmap for future research. *J Virol* 98:e0043724. <https://doi.org/10.1128/jvi.00437-24>
84. Arduino PG, Porter SR. 2008. Herpes simplex virus type 1 infection: overview on relevant clinico-pathological features. *J Oral Pathol Med* 37:107–121. <https://doi.org/10.1111/j.1600-0714.2007.00586.x>
85. Pintó RM, Pérez-Rodríguez F-J, Costafreda M-I, Chavarria-Miró G, Guix S, Ribes E, Bosch A. 2021. Pathogenicity and virulence of hepatitis A virus. *Virulence* 12:1174–1185. <https://doi.org/10.1080/21505594.2021.1910442>
86. Romero JR, Selvarangan R. 2011. The human parechoviruses: an overview. *Adv Pediatr* 58:65–85. <https://doi.org/10.1016/j.yapd.2011.03.008>
87. Reuter G, Pankovics P, Boros Á. 2017. Saliviruses—the first knowledge about a newly discovered human picornavirus. *Rev Med Virol* 27:27. <https://doi.org/10.1002/rmv.1904>
88. Robinson CM, Seto D, Jones MS, Dyer DW, Chodosh J. 2011. Molecular evolution of human species D adenoviruses. *Infect Genet Evol* 11:1208–1217. <https://doi.org/10.1016/j.meegid.2011.04.031>
89. Robertson BH, Jia XY, Tian H, Margolis HS, Summers DF, Ehrenfeld E. 1992. Serological approaches to distinguish immune response to hepatitis A vaccine and natural infection. *Vaccine* 10 Suppl 1:S106–S109. [https://doi.org/10.1016/0264-410x\(92\)90559-3](https://doi.org/10.1016/0264-410x(92)90559-3)
90. Tokarz R, Haq S, Sameroff S, Howie SRC, Lipkin WI. 2013. Genomic analysis of coxsackieviruses A1, A19, A22, enteroviruses 113 and 104: viruses representing two clades with distinct tropism within enterovirus C. *J Gen Virol* 94:1995–2004. <https://doi.org/10.1099/vir.0.053462-0>
91. Brouwer L, Benschop KSM, Nguyen D, Kamau E, Pajkrt D, Simmonds P, Wolthers KC. 2020. Recombination analysis of non-poliovirus members of the enterovirus C species; restriction of recombination events to members of the same 3DPol cluster. *Viruses* 12:706. <https://doi.org/10.3390/v12070706>
92. Basit H, Tyagi I, Koirala J. 2023. Hepatitis CStatPearls. *StatPearls Publishing, Treasure Island (FL)*.
93. Deeks SG, Overbaugh J, Phillips A, Buchbinder S. 2015. HIV infection. *Nat Rev Dis Primers* 1:15035. <https://doi.org/10.1038/nrdp.2015.35>
94. Fink ZW, Martinez V, Altin J, Ladner JT. 2020. PepSIRF: a flexible and comprehensive tool for the analysis of data from highly-multiplexed DNA-barcoded peptide assays. *arXiv:2007*.
95. Brown AM, Bolyen E, Raspet I, Altin JA, Ladner JT. 2022. PepSIRF + QIIME 2: software tools for automated, reproducible analysis of highly-multiplexed serology data. <https://doi.org/10.48550/ARXIV.2207.11509>
96. Mina MJ, Kula T, Leng Y, Li M, de Vries RD, Knip M, Siljander H, Rewers M, Choy DF, Wilson MS, Larman HB, Nelson AN, Griffin DE, de Swart RL, Elledge SJ. 2019. Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science* 366:599–606. <https://doi.org/10.1126/science.aay6485>
97. Seabold S, Perktold J. 2010. Statsmodels: econometric and statistical modeling with Python. Python in Science Conference; Austin, Texas. *SciPy*. <https://doi.org/10.25080/Majora-92bf1922-011>
98. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
99. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35:4453–4455. <https://doi.org/10.1093/bioinformatics/btz305>
100. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29. <https://doi.org/10.1186/1471-2148-6-29>