

1 Mapping disparities in viral infection rates using highly-multiplexed serology

2
3 Alejandra Piña*¹, Evan A Elko*¹, Rachel Caballero², Mary Mulrow², Dan Quan^{2,3,4}, Lora
4 Nordstrom², John A Altin⁵, Jason T Ladner^{1#}

5
6 Affiliations:

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

8 ² Valleywise Health, Phoenix, AZ, USA

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

10 ⁴ Creighton University, School of Medicine, Phoenix, AZ, USA

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

12 * These authors contributed equally

13 # Correspondence: Jason.Ladner@nau.edu

14

15 Abstract

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

32

33 Importance

34 Our understanding of population-level virus infection rates and associated health disparities is
35 incomplete. In part, this is because of the high diversity of human-infecting viruses and the
36 limited breadth and sensitivity of traditional approaches for detecting infection events. Here, we
37 demonstrate the potential for modern, highly-multiplexed antibody detection methods to greatly
38 increase our understanding of disparities in rates of infection across subpopulations (e.g.,
39 different sexes or ethnic groups). The use of antibodies as biomarkers allows us to detect
40 evidence of past infections over an extended period of time, and our approach for highly-
41 multiplexed serology (PepSeq) allows us to measure antibody responses against 100s of viruses
42 in an efficient and cost-effective manner.

43

44

45 **Introduction**

46 Despite advancements in the development of medical interventions, the global burden of
47 disease caused by viral pathogens remains substantial and highly diverse. This burden includes a
48 wide range of morbidities associated with active viral replication ranging in severity from fever,
49 muscle aches, and rash to encephalitis, immunosuppression, respiratory failure, and congenital
50 birth defects (1–4). Additionally, it includes an array of clinical sequelae of infection (e.g.,
51 Guillain-Barre syndrome, Multisystem inflammatory syndrome, Long COVID), many of which
52 remain poorly understood (5–7). Viral infections have even been linked to the onset of a number
53 of non-communicable diseases, such as myocarditis (8), diabetes (9), celiac disease (10), obesity
54 (11), multiple sclerosis (12, 13), cancer (14), and Alzheimer's disease (15).

55 Although the health effects caused by viral infections are observed widely in the general
56 population, currently documented national trends highlight several sex- and ethnicity-based health
57 disparities in the prevalence of viral infections and thus, virus-associated disease. For example, in
58 the United States (US), seroprevalence of human papillomavirus (HPV) and herpes simplex virus
59 2 (HSV-2) have been shown to be roughly twice as high among women compared to men (16, 17).
60 In contrast, human immunodeficiency virus 1 (HIV-1) disproportionately affects men in the US,
61 particularly men who have sex with men. In 2021, 69% of new HIV-1 diagnoses in the US were
62 among men who have sex with men, despite this group representing only 3.9% of the US
63 population (18, 19). Ethnic disparities have also been documented for HIV-1 in the US. A 2015-
64 2019 CDC surveillance report showed the incidence of HIV-1 infection among Hispanics was four
65 times higher than that among non-Hispanic Whites. Furthermore, disparities in viral infections can
66 vary in space and time. For example, over a 12 year period the average annual incidence per
67 100,000 people of hepatitis A virus (HAV) in a Native American population in Arizona decreased
68 from 289 to 6 due to vaccination efforts (20). Vector-borne viruses provide striking examples of
69 geographical disparities. For example, dengue virus is transmitted by mosquitoes of the genus
70 *Aedes* that live primarily in tropical regions (21). In 2020 the number of locally transmitted cases
71 of dengue virus was 0.025 per 100,000 people in the continental United States, compared to 23.68
72 cases per 100,000 people in Puerto Rico (22).

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

83 Recent advances in serological methods have overcome previous limitations in breadth and
84 are enabling unprecedented views into the viral exposure histories of individuals (25–28). Using

85 current approaches for highly-multiplexed serology (e.g., PepSeq (29), PhIP-Seq (30)) it is now
86 possible to characterize antibody binding to 100,000s of antigens in a single assay using <1 μ L of
87 blood. In this study, we utilize the PepSeq platform to demonstrate the potential of highly-
88 multiplexed serology to broadly characterize differences in seroprevalence across the human
89 virome between different demographic groups. Through the characterization of antiviral antibodies
90 in samples collected over ~1 month within a single healthcare system in Phoenix, AZ, we
91 document significant differences in infection rates between the Hispanic White (HW) and non-
92 Hispanic White (NHW) populations for several viruses, including some that are rarely included in
93 population-level surveys.

94

95 **Results**

96

97 **Study population**

98 Between late May and early June 2020, we collected 400 remnant serum samples from 11
99 Valleywise Health facilities, a large safety net hospital system in Phoenix, AZ. These samples
100 were distributed equally among four subpopulations: 100 HW men, 100 HW women, 100 NHW
101 men and 100 NHW women. To minimize the impact of age-related differences in seropositivity,
102 we limited our focal age range to 30-60 years, and the age distributions were not significantly
103 different (t-test) between genders or ethnicities, with mean age ranging between 45.1 and 46.2
104 (Figure 1A). We also investigated payor source for each sampled individual (Figure 1B), as this
105 can serve as a proxy for socioeconomic status (31). Overall, the vast majority of the individuals
106 included in our study were either covered by a government insurance plan (52.8%; Tricare,
107 Medicare and/or Medicaid) or were uninsured (31.3%; self-pay). Only ~13.1% of the individuals
108 in our study were covered by commercial insurance plans and this percentage did not vary
109 considerably among our subpopulations, though we observed a slightly higher rate of commercial
110 payor for NHW females (20.2%) compared to the other three groups (10-11.2%). However, we
111 did observe substantial differences among our subpopulations in the proportions covered by either
112 government programs or uninsured. The NHW populations were covered in higher proportions by
113 government insurance programs (HW-M: 44.9%, HW-F: 16%, NHW-M: 79.8%, NHW-F: 72.7%),
114 while the HW populations were more likely to be uninsured (HW-M: 39.8%, HW-F: 72%, NHW-
115 M: 10.1%, NHW-F: 3%).

116

117 **PepSeq Analysis**

118 All 400 serum samples were assayed, in duplicate, using our human virome version 1
119 (HV1) PepSeq library (26), and we obtained an average of 2.2M Illumina sequencing reads per
120 sample, which equates to an average of 9.2 reads per unique HV1 peptide per sample. Four samples
121 were excluded from further analysis due to lower than normal correlation between replicates,
122 which may indicate the occurrence of molecular bottlenecks or contamination of one of the
123 replicates while performing the assay (29). These included two HW males, one NHW male and
124 one NHW female. Therefore, all the analyses presented here include a total sample size of 396.

125 Based on visual comparisons of experimental samples and buffer-only negative controls
126 (Supplemental Figure 1), as well the analysis of a separate group of negative controls that were
127 not considered in the formation of bins or normalization of the data, we chose a set of four different
128 Z score thresholds (10, 15, 20 and 25) for identifying enriched peptides. Higher thresholds are
129 expected to have reduced sensitivity, but increased specificity. To estimate the false positive rate
130 at each of these thresholds, we analyzed all pairwise combinations of 9 buffer-only negative
131 controls (n=36). We observed an average of 5.5, 1.1, 0.28 and 0.17 putatively enriched peptides
132 from these control pseudoreplicate analyses for thresholds of 10, 15, 20 and 25, respectively. In
133 contrast, from the assays of serum samples, we observed an average of 1335, 1071, 926 and 828
134 enriched peptides for thresholds of 10, 15, 20 and 25, respectively. This equates to expected false
135 positive rates of approximately 0.41%, 0.1%, 0.03% and 0.02%, respectively.

136 To broadly characterize enrichment patterns within our data set, we averaged the number
137 of enriched peptides across all Z score thresholds for each sample. We did not observe significant
138 differences between the average number of enriched peptides by gender (F=1009.92, M=1070.30;
139 t-test p-value=0.135) or ethnicity (HW=1044.37, NHW=1035.55; t-test p-value=0.827) (Figure
140 2A). We also observed no significant correlation between the number of enriched peptides and age
141 (Figure 2B; Pearson correlation p-value=0.221).

142 The enriched peptides for each Z score threshold were then converted into putative virus
143 species-level serostatus calls using two different sets of potential viruses: one included all 390
144 HIV target species as viruses to which the study individuals may have been exposed (“Full”),
145 while the other only included 93 viruses (“Focal93”), excluding those that are unlikely to have
146 been encountered by our focal population (Table S1). Overall, in comparing the results from these
147 two different sets of potential viruses, we observed strong positive correlations in individual-level
148 estimates of the number of seropositive species (Pearson’s correlation coefficients: 0.957 - 0.963;
149 p-values: $1.13e^{-214}$ - $2.50e^{-226}$) and in species-level estimates of seroprevalence (Pearson’s
150 correlation coefficients: 0.983 - 0.997; p-values: $7.35e^{-60}$ - $2.69e^{-87}$). As expected, we also found
151 that most of the seropositive calls using the Full species set were for viruses included in our
152 Focal93 list (76%, 82%, 86%, and 89% for Z score thresholds of 10, 15, 20, and 25, respectively),
153 despite the fact that the Focal93 viruses make up a minority of the Full list of virus species (24%).
154 The relatively small number of predicted exposures to viruses outside of our Focal93 subset are
155 likely due to the presence of cross-reactive antibodies and/or true exposure to unusual or
156 uncharacterized viruses. Separately for each set of viruses, we averaged the number of species per
157 sample across all Z score thresholds to broadly characterize patterns within our subpopulations,
158 and we found no significant differences (t-test) in the average number of predicted seropositive
159 species by gender (Focal93: F=29.44, M= 29.92; Full: F=33.58, M=34.11) or by ethnicity
160 (Focal93: HW=29.84, NHW=29.52; Full: HW=34.20, NHW=33.49) (Figure 3A). However, we
161 did observe a significant positive correlation between age and the number of predicted seropositive
162 species from both sets of potential viruses (Focal93: Pearson correlation p-value=0.007, Figure
163 3B; Full: Pearson correlation p-value=0.001, not shown).

164 Next, to assess the performance of PepSeq in determining population level seroprevalence
165 across a wide range of virus species, we calculated the estimated seroprevalence for the full cohort
166 of 396 samples. This analysis resulted in the estimated seroprevalence for 306 virus species.
167 Among the 20 virus species with the highest seroprevalence, the *Herpesviridae* and *Picornaviridae*
168 families were the most frequently represented (Figure 3C). Seven virus species had an estimated
169 seroprevalence greater than 0.9 (90%; Human gammaherpesvirus 4, Rhinovirus A-C, Human
170 orthopneumovirus, Norwalk virus and Human respirovirus 3) (Figure 3C). For 12 virus species
171 with published seroprevalence studies in the United States, we then compared the PepSeq
172 estimates of seroprevalence with the published estimates based on more traditional singleplex
173 assays (32–41). There was a highly significant correlation between the PepSeq estimated
174 seroprevalence and the estimated seroprevalence found in the literature (p -value=2.72e-5, Pearson
175 $R=0.917$) (Figure 3D).

176

177 **Identification of disparities**

178 To identify statistically significant differences in estimated seroprevalence among our
179 subpopulations, we fit a binomial generalized linear model (GLM) with a single dependent
180 variable (seroprevalence) and three independent variables (ethnicity, gender, and age). The results
181 of this analysis using the Full and Focal93 virus sets were remarkably consistent, and all the viruses
182 that exhibited significant differences between subpopulations were present within the Focal93
183 subset (Table S2). Therefore, we only present the results for the Focal93 set of viruses.

184 In general, across virus species, we observed higher total seropositivity in older individuals,
185 but there was no consistent directional change associated with gender or ethnicity (Supplemental
186 Figure 2). However, no individual viruses exhibited significant correlations between age and
187 seropositivity after correcting for multiple tests, which may be related to the limited range of ages
188 included in this study (30-60 years old). We also found no significant differences in estimated
189 seroprevalence between genders. However, we did observe several viruses with significant
190 correlations between seropositivity and ethnicity, and these patterns were generally consistent
191 across our four Z score thresholds (Figure 4A).

192 In total, ten virus species exhibited ethnicity p -values < 0.05 across all four Z score
193 thresholds: cytomegalovirus (CMV, aka human herpesvirus 5), herpes simplex virus 1 (HSV-1)
194 and 2 (HSV-2), human herpesvirus 7 (HHV-7), hepatitis A virus (HAV), salivirus A (SaV-A),
195 parechovirus A (PeV-A), *Enterovirus C* (EV-C), human adenovirus D (HAdV-D) and human
196 coronavirus OC43 (HCoV-OC43). After Bonferroni correction for multiple tests, seven of these
197 remained significant at ≥ 1 Z score threshold and five remained significant across all four Z score
198 thresholds (Figure 4A). Within the HW subpopulation, we observed significantly higher
199 seropositivity for CMV, HSV-1, HAdV-D and HAV across all four Z thresholds and for SaV-A at
200 a Z threshold of 10 (Figure 4A). Within the NHW subpopulation, we observed significantly higher
201 seropositivity for HHV-7 across all thresholds and for PeV-A at Z thresholds of 15 and 20.

202 For all ten viruses showing significant or nearly significant differences in seroprevalence
203 between ethnicities, we also looked at the relationship between seropositivity and insurance status.

204 Specifically, using a Z score threshold of 15, we compared seroprevalence estimates among three
205 subsets of our study population: insured NHWs (n=185), insured HWs (n=87), and uninsured HWs
206 (n=111). Uninsured NHWs were excluded because of a low sample size (n=13), but generally
207 showed seroprevalence estimates similar to insured NHWs. Our results showed that for most
208 viruses, estimated seroprevalence within insured HWs was intermediate between the NHW insured
209 and HW uninsured subpopulations (Figure 4B). In fact, when considering all ten viruses together,
210 we observed a significant difference between the normalized estimated seroprevalences in the
211 insured HW group compared to the uninsured HW group (absolute values, paired t-test p-
212 value=0.013). For all six of the viruses with higher seropositivity among HWs, seropositivity was
213 highest among uninsured individuals, and for half (2/4) of the viruses with lower seropositivity in
214 HWs, seropositivity was lowest in the HW uninsured group. Two viruses, HHV-7 and HCoV-
215 OC43, which both had lower seropositivities among HWs, do not follow this trend and have
216 slightly higher seroprevalences in the uninsured HWs compared to the insured HWs.

217 Vaccines are available for two of the virus species with higher seroprevalence in our HW
218 population (HAV and EV-C, which includes poliovirus), and therefore the differences we observed
219 could be due to differences in rates of either natural infection or vaccination. Therefore, we further
220 dissected the antibody responses against these viruses by examining protein- and peptide-level
221 reactivity profiles. Notably, all HAV vaccines approved for use in the US are inactivated and it
222 has been shown that antibody responses to natural infection and vaccination can be differentiated
223 by measuring the response to nonstructural proteins, to which a response will only be generated
224 with a natural infection (i.e., when there is virus replication and production of non-structural
225 proteins) (42). Therefore, to examine the role of natural exposure to HAV in the observed
226 difference in seropositivity, we mapped all enriched HAV peptides for each seropositive sample
227 across the HAV proteome. We did not observe a significant difference between HW and NHW
228 cohorts in the proportion of seropositive individuals with ≥ 1 enriched peptide from an HAV non-
229 structural protein (Fisher's exact test p-value=0.77). In fact, we observed high rates of reactivity
230 against non-structural HAV proteins across both groups (HW=53/71, NHW=17/21) (Figure 5A).
231 These findings indicate that most of the seropositive individuals in our study have likely been
232 naturally infected by HAV and that the observed difference in seroprevalence between HW and
233 NHW is probably not driven by differences in vaccination rate.

234 Next, we sought to determine the role of poliovirus vaccination on the increased
235 seropositivity to EV-C in HWs compared to NHWs. In this case, both inactivated and attenuated
236 (replication competent) vaccines were used in the US prior to 2000, and both are still administered
237 in Mexico, which borders Arizona and is the most common source of immigrants in the state (43).
238 Therefore, instead of comparing protein-level patterns of antibody reactivity, we compared
239 antibody reactivity to peptides designed specifically from the three strains of poliovirus.
240 Specifically, we reran our estimates of seropositivity after separating the three polioviruses
241 (UniProt Accessions can be found in Table S3) from the rest of EV-C. In this new analysis, we
242 saw no significant difference in estimated serostatus for poliovirus (p-value = 0.53) with 78 (39%)
243 HW and 83 (42%) NHW positive samples (Figure 5B). However, we did observe a highly

244 significant difference in seropositivity between HWs and NHWs for “Other EV-C” strains, with
245 94 (47%) and 14 (7%) seropositive samples, respectively (p -value <0.0001) (Figure 5B). These
246 results indicate that the observed disparity in EV-C seropositivity between HWs and NHWs is not
247 driven by differences in natural infection or vaccination with poliovirus but is likely caused by
248 differences in infection rate with other EV-C viruses.

249 To determine whether there were particular non-polio EV-C viruses that were driving the
250 disparity seen in the “Other EV-C” group, we further analyzed the reactive peptides assigned to
251 this group. First, we split the 27 International Committee on Taxonomy of Viruses (ICTV) listed
252 EV-C isolates into six phylogenetic clades (Figure 6A, one clade includes only the polioviruses).
253 Next, we assigned each “Other EV-C” peptide to the most similar EV-C clade (based on shared
254 amino acid 7mers) and assigned these peptides scores equivalent to the number of contained 7mers
255 that are unique to that clade. We then calculated relative peptide scores for each EV-C clade by
256 summing the scores for 1) all “Other EV-C” peptides in the full HV1 PepSeq library (null
257 distribution) and 2) the subsets of enriched “Other EV-C” peptides observed in the HW and NHW
258 samples that were seropositive for “Other EV-C” (Figure 6B, Supplemental Figure 3). We
259 observed antibody reactivity to peptides assigned to all five “Other EV-C” clades, and the clade-
260 specific relative peptide scores varied substantially between individuals (Supplemental Figure 3).
261 These results suggest that a variety of different EV-C viruses may be contributing to the observed
262 disparity in EV-C seropositivity between HWs and NHWs. However, some of these clades (e.g.,
263 EV-C_2 and EV-C_3) may be more common than others, based on differences in the relative
264 peptide scores between the expected (“Full Library”) and observed (“NHW”, “HW”) distributions
265 (Figure 6B).

266

267 **Discussion**

268 In this study, we used PepSeq, a highly-multiplexed serology platform, to broadly assess
269 virus infection histories and identify differences in seropositivity among various subsets of the
270 population served by Valleywise Health in Phoenix, AZ. PepSeq allows for 100,000s of peptide
271 antigens to be simultaneously assayed for antibody reactivity, and thus, it has the potential to
272 facilitate the comprehensive characterization of differences in viral infection rates among subsets
273 of a community. In contrast, the singleplex nature of traditional serological techniques (e.g.,
274 ELISA) has required that previous studies focus on a small number of high priority viruses (44,
275 45).

276 In general, our results were consistent with expectations from previously published studies.
277 First, we observed a general trend toward higher seropositivity with increasing age, a pattern that
278 has been reported for a variety of viruses (32–34, 39). In fact, we observed this pattern at two
279 levels: 1) a positive correlation between an individual’s age and the number of seropositive virus
280 species called (Figure 3B) and 2) negative age-associated GLM coefficients for most viruses,
281 indicative of higher seroprevalence with increased age (Supplemental Figure 2). Second, we found
282 that overall estimates of seroprevalence for individual viruses were broadly consistent with
283 expectations from molecular and singleplex serological surveys. For example, with a Z score

284 threshold of 15, seven viruses had estimated seropositivities $\geq 90\%$ (Figure 3C). Among these were
285 five common respiratory viruses, for which the expected seroprevalence is near 100% by
286 adulthood (46): human rhinoviruses A, B and C, human orthopneumovirus (aka respiratory
287 syncytial virus) (47) and human respirovirus 3 (aka human parainfluenza virus 3) (48). Also
288 included in this set are human gammaherpesvirus 4 (aka Epstein-Barr virus) and Norwalk virus
289 (aka norovirus), consistent with published serological surveys of adults in the US (33, 38). We
290 also found that relative seroprevalence estimates for closely related viruses were generally
291 consistent with documented differences in abundance. For example, while we estimated a
292 seroprevalence for the more common human respirovirus 3 of 91%, the less common
293 “parainfluenza” viruses (human respirovirus 1 and rubulaviruses 2 and 4) were estimated to have
294 seroprevalences of 25-36% (49). Similarly, we estimated the seroprevalence of HSV-1 (85%) to
295 be $\sim 2.2x$ higher than that of HSV-2 (38%) (Figure 3D), reflecting known differences in prevalence
296 of these related viruses (50). Taken together, these observations provide strong support for the use
297 of our highly-multiplexed serology approach for broadly characterizing virus infection histories.

298 Although we did not observe any statistically significant differences in seropositivity
299 between males and females in our study, we did observe significant differences between HWs and
300 NHWs for seven different viruses, and these differences were largely consistent across different Z
301 score thresholds for peptide enrichment (Figure 4A). For five of these viruses (CMV, HSV-1,
302 HAV, SaV-A and HAdV-D), we observed significantly higher seropositivity within our HW
303 subpopulation, while the other two (HHV-7 and PeV-A) showed higher seropositivity in our NHW
304 subpopulation, and several of these significant trends are consistent with published studies, lending
305 additional credibility to our highly-multiplexed approach. For example, we detected higher
306 seropositivity among HWs for both CMV (10-20% higher than among NHWs) and HSV-1 (10-
307 30% higher), consistent with the results of previous US population studies that used singleplex
308 antibody assays (32, 50). We also observed significantly higher seropositivity for HAV among
309 HWs, which is consistent with published rates of seropositivity in the US, as well as known
310 differences in seropositivity between the US and Mexico (34, 51–53).

311 We also observed several previously undocumented disparities in infection history between
312 our HW and NHW populations, most of which involved viruses that are rarely targeted in
313 serosurveys. One example is the recently described picornavirus SaV-A, which had $\sim 10\%$ higher
314 seropositivity within our HW subpopulation. Little is known about the seroprevalence of SaV-A
315 in the US, however, it has been documented in wastewater samples from multiple US states, and
316 a study in Arizona found evidence of SaV-A in 15% of wastewater samples (54, 55). Another
317 previously undocumented disparity involves HAdV-D, for which we observed 20-30% higher
318 seropositivity within our HW subpopulation (Figure 4). HAdV-D is a highly diverse species that
319 has been shown to cause severe disease such as epidemic keratoconjunctivitis in
320 immunocompromised individuals. However, little is known about the general health impact of
321 HAdV-D, and therefore, it is difficult to deduce the impact this virus might have on the wellbeing
322 of this population (56). Finally, we estimated significantly higher seroprevalence in our NHW
323 subpopulation for two understudied viruses: PeV-A and HHV-7. PeV-A is a widespread virus that

324 is associated with respiratory and gastrointestinal symptoms (57–59). However, certain strains of
325 PeV-A are associated with more severe disease, such as meningitis and sepsis-like illness in infants
326 (58). HHV-7 is a highly prevalent virus, which is primarily contracted in early childhood, resulting
327 in life-long latent infections that typically remain asymptomatic. However, HHV-7 has also been
328 linked with febrile seizures and as a possible cause of encephalitis (60–62). More work is needed
329 to understand the full impact that these undocumented disparities may have on human health
330 within these populations.

331 Arizona is home to a large and diverse immigrant population and this has likely contributed
332 to our observed differences in seropositivity among ethnicities. According to estimates from 2018,
333 immigrants (i.e., foreign-born individuals) comprised 13% of the population in Arizona, 16% of
334 the state’s population were native-born Americans with at least one immigrant parent, and 55% of
335 all immigrants in Arizona were from Mexico, with the next most common countries of origin
336 (Canada, India, Philippines) each accounting for only 4% of the immigrant population (63).
337 Although we did not have access to the immigration status for the individuals included in this
338 study, we were able to examine the payor source associated with each individual’s medical visit,
339 and we used a lack of insurance coverage (i.e., “self pay”) as a proxy for immigration status. In
340 other words, it was assumed that the uninsured population would contain a higher proportion of
341 immigrants compared to the insured population, and consistent with this assumption, we observed
342 a much higher proportion of uninsured individuals within our HW subpopulation compared to our
343 NHW subpopulation (Figure 1B).

344 By comparing seropositivity estimates between the uninsured and insured HW
345 subpopulations, we were able to demonstrate that, for the majority of viruses that exhibited
346 significant or near significant differences in seropositivity between ethnicities, the insured HW
347 population represented an intermediate level of seropositivity between NHWs and uninsured HWs
348 (Figure 4B). This pattern is consistent with the hypothesis that a substantial portion of the
349 differences in seroprevalence between ethnicities can be attributed to differences in seroprevalence
350 between immigrant and non-immigrant populations. Antibody responses to viruses can be long-
351 lived. Therefore, one possible explanation is that we are seeing evidence of differences in virus
352 exposure rates, possibly during adolescence, for those who were raised in the US versus those who
353 were raised outside of the US, potentially in less developed, resource-limited countries. Many of
354 the viruses flagged by our analysis are commonly encountered during childhood, and several are
355 known to be more prevalent outside of the US (CMV, HSV-1, and HAV) (64–66). However, other
356 factors, such as differences in the dynamics of virus transmission in immigrant communities within
357 the US could also be contributing to the observed patterns.

358 Virus transmission mechanisms may also play an important role in determining which
359 viruses are most likely to be associated with disparities among ethnic groups. Among the seven
360 viruses with a significant difference in seroprevalence between HWs and NHWs, three are
361 primarily transmitted through intimate contact (CMV, HHV-7, HSV-1), three through the fecal-
362 oral route (HAV, PeV-A, SaV-A) and one (HAdV-D) has been associated with a variety of
363 transmission routes, including close personal contact, the fecal-oral route, and respiratory droplets.

364 Notably absent from this list is respiratory viruses that are primarily spread through airborne
365 transmission, which make up roughly 23% of the focal93 viruses (Table S4). This absence of
366 airborne transmitted viruses is likely due to the wider potential radius for spread from person to
367 person and suggests that viruses that rely on close contact for successful transmission are more
368 likely to be associated with population-level disparities.

369 Among our differentially seropositive species, two (HAV and EV-C) include viruses for
370 which there are widely available vaccines, and for both of these, our highly-multiplexed assay
371 allowed us to assess the relative roles of vaccination and natural infection on the observed
372 differences in seropositivity. For HAV, we were able to differentiate antibody reactivity from
373 vaccination and natural infection by leveraging the ability of PepSeq to simultaneously measure
374 antibody reactivity across multiple HAV protein targets. Specifically, we compared patterns of
375 antibody reactivity between the structural and non-structural proteins of HAV (Figure 5A). For
376 nonstructural proteins to be produced, active replication of viral particles must occur (67).
377 However, all available vaccines in the US contain inactivated viruses, and therefore, vaccination
378 will not elicit antibodies that target nonstructural proteins. Our results show high levels of antibody
379 reactivity against nonstructural HAV proteins in both HW and NHW individuals. This suggests
380 that the higher seroprevalence among HWs is not because of higher vaccination rates in this
381 population.

382 There is also a commonly used vaccine that includes three viruses that belong to the EV-C
383 species — poliovirus 1, 2, and 3 — and both inactivated and live attenuated versions were in use
384 during the lifetimes of our participants. Therefore, to control for differences in antibody responses
385 that may be driven by differences in rates of vaccination, we separated our EV-C peptides into two
386 categories: those that share at least one 7mer with any of the three polioviruses (i.e., those most
387 similar to vaccine antigens) and those that don't. We saw no difference between our ethnicities in
388 reactivity against the peptides most likely to be recognized by vaccine-induced antibodies, but a
389 highly significant difference in reactivity against “Other EV-C” peptides (40% higher
390 seropositivity among HWs; Figure 5B). Further analysis showed that the observed antibody
391 reactivity profiles are consistent with past exposures to a wide variety of EV-C viruses
392 (Supplemental Figure 3), but that some phylogenetic clades may be contributing more than others
393 to the observed disparity in EV-C infections between HWs and NHWs (Figure 6B). Notably, while
394 EV-C_5 peptides are overrepresented in our starting library, they are comparably underrepresented
395 among our enriched peptides, especially in our HW subpopulation (Figure 6B). Interestingly, EV-
396 C_5 has been predominantly isolated from nasal/throat swabs and nasopharyngeal aspirates,
397 suggesting these viruses are likely transmitted via respiratory droplets (68, 69). In contrast, the
398 other EV-C clades have been isolated almost universally from stool samples, indicating that they
399 primarily infect the gastrointestinal tract and are likely to be spread through the fecal-oral
400 transmission pathway (68, 69). This tissue specific tropism aligns with our general hypothesis that
401 the population level virus infection disparities are more commonly driven by viruses that require
402 close contact transmission routes.

403

404 **Conclusion**

405 Overall, our highly-multiplexed serology assay was successful in broadly characterizing
406 antiviral antibody reactivities and allowed us to infer individual infection histories across the
407 virome. The recapitulation of several known differences in seropositivity between two ethnic
408 groups provides confidence in the quality of our analysis and highlights promising future
409 applications for this type of approach. Additionally, our study revealed several previously
410 undocumented disparities in virus seropositivity between HWs and NHWs in our study population.
411 Future studies will be needed to better understand the clinical significance of these differences in
412 infection rates, and to develop medical and social interventions to minimize the impact of these
413 disparities. Our results also demonstrate the potential for highly-multiplexed serology to finely
414 dissect the specificity and breadth of antibody responses, thus enabling an unprecedented view
415 into an individual's history of infection.

416

417 **Materials and Methods**

418 Study population and sample collection

419 In total, 400 serum samples were obtained from Valleywise Health in Phoenix, AZ. These
420 samples were collected in late May and early June 2020. They were remnant samples initially
421 collected as a part of the patients' standard of care, and were collected from several different
422 facilities and in a variety of contexts including outpatient encounters (55.5%), inpatient encounters
423 (35%) and emergency department visits (9.5%). Researchers at Northern Arizona University
424 (NAU) did not have access to any identifiable patient information. This study was reviewed and
425 approved by the Valleywise Health and NAU Institutional Review Boards (approval number
426 1545420).

427 To maximize statistical power to detect differences in seroprevalence, our cohort for this
428 study was equally divided among four sub-populations: HW males (n=100), HW females (n=100),
429 NHW males (n=100) and NHW females (n=100). To minimize the effect of age in detecting
430 differences in seroprevalence, we selected only individuals within an age range of 30-60 years old.
431 Self-reported ethnicity, gender and age were the only characteristics considered for inclusion in
432 this study. However, because Valleywise Health is a safety net hospital, we do not expect our study
433 population to represent a random sampling of the population of Phoenix, AZ. Rather, it is likely to
434 include a higher proportion of individuals with low income and from under-served populations.
435 There is also some potential for bias associated with the use of remnant samples collected from
436 patients actively receiving medical care (compared to a random sample of adults). However, given
437 the wide variety of encounter types represented in our sample, we expect any impact to be minimal.

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

443 For each patient, we also obtained payor source, as this can serve as a useful, though
444 incomplete, indicator of socio-economic status. As the exact payor source is highly variable among
445 individuals, we reduced the complexity of this categorical variable by assigning every individual
446 to one of six general categories: 1) “Commercial”, which included all commercial health plans; 2)
447 “Medicaid”, which included both Arizona Health Care Cost Containment plans and out-of-state
448 Medicaid; 3) “Medicare”; 4) “Dual-SNP”, which included any dual special needs plans for
449 individuals who qualify for both Medicaid and Medicare; 5) “Self Pay”, for individuals without
450 insurance and 6) “Other”, which served as a final catch-all category that included funding through
451 charitable organizations like the Ryan White HIV/AIDS Program and other government plans such
452 as Tricare (Table S5).

453

454 PepSeq Library Design and Assay

455 To broadly assess antiviral antibody reactivity, we utilized the PepSeq platform to perform
456 highly-multiplexed peptide-based serology (29). Specifically, we used the human virome version
457 1 (HV1) library described in (26). In brief, the HV1 library consists of 244,000 unique DNA-
458 peptide conjugates (i.e., PepSeq probes). The variable peptide portion of each molecule is 30
459 amino acids long and the peptides were designed to broadly cover potential linear epitopes present
460 in the proteins of viruses known to infect humans. Libraries of these PepSeq probes are created
461 through a series of bulk, *in vitro* enzymatic reactions (29).

462 Each assay was conducted as described in Ladner et al. (26). Broadly, the PepSeq assay
463 involves the incubation of serum with a diverse pool of PepSeq probes. Immunoglobulin G (IgG)
464 is then precipitated using magnetic protein G beads, non-binding PepSeq probes are washed away,
465 and the relative abundance of each probe is quantified using PCR and high-throughput sequencing
466 of the DNA portion of the molecules (29). Specifically, 5 μ L of a 1:10 dilution of serum in
467 Superblock T20 (Thermo) was added to 0.1 pmol of the PepSeq library for a total volume of 10
468 μ L and was incubated at 20°C overnight. The binding reaction was incubated with pre-washed
469 protein G-coated beads (Thermo) for 15 minutes, after which the beads were hand washed 11 times
470 with 1x PBST. After the final wash, beads were resuspended in 30 μ L of water and heated to 95°C
471 for 5 minutes to elute the bound PepSeq probes. Elutions were amplified and indexed using
472 barcoded DNA oligonucleotides (Table S6). Following PCR, a standard bead cleanup was
473 performed and products were individually quantified (Quant-It, Thermo Fisher), pooled, re-
474 quantified (KAPA Library Quantification Kit, Roche) and sequenced on a NextSeq instrument
475 (Illumina). For this study, each sample was assayed in duplicate and ≥ 1 buffer-only negative
476 controls were included on each 96-well assay plate. Potential batch effects were controlled through
477 equal representation of each of our four focal subpopulations on each plate, as well as through the
478 inclusion of negative controls from all plates in read count normalization and the generation of the
479 peptide bins.

480

481

482

483 PepSeq Analysis

484 We used PepSIRF v1.5.0 (70, 71) to analyze the high-throughput sequencing data.
485 Demultiplexing and assignment of reads to peptides was done using the *demux* module of PepSIRF
486 allowing up to 1 mismatch within each of the index sequences (12 and 8 nt, respectively) and up
487 to 3 mismatches with the expected DNA tag (90 nt). Z scores were calculated using the *zscore*
488 module of PepSIRF, which implements a method adapted from (72). This process involved the
489 generation of peptide bins, each of which contained ≥ 300 peptides with similar expected
490 abundances in our PepSeq library. Expected abundance for each peptide was estimated using
491 buffer-only negative controls. In total, 12 independent buffer-only controls from 7 different assay
492 plates were used to generate the bins for this study. The raw read counts from each of these controls
493 were first normalized to reads per million (RPM) using the column sum normalization method in
494 the *norm* module of PepSIRF. This served to normalize for differences in total sequencing depth
495 between samples. Bins were then generated using the *bin* PepSIRF module. RPM counts for each
496 peptide were then further normalized by subtracting the average RPM count observed within our
497 buffer-only controls. This second normalization step was used to control for any differences in
498 initial relative abundance among peptides contained within the same bin. Each Z score was
499 calculated using peptides contained within the same bin and corresponds to the number of standard
500 deviations away from the mean, with the mean and standard deviation calculated using the 95%
501 highest density interval to exclude any enriched peptides.

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

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

521 The HV1 PepSeq library includes peptides derived from 390 virus species, but many of
522 these viruses have only very rarely been associated with human infections (e.g., foot and mouth

523 disease virus, simian foamy virus) and/or occur in distant and spatially restricted parts of the world
524 (e.g., Ebola virus, Crimean-Congo hemorrhagic fever virus), and therefore it is very unlikely that
525 our focal population would have been exposed to these viruses. To determine whether the inclusion
526 of these viruses was impacting the results of our analysis, we generated two distinct linkage maps.
527 The first (“Full”, Table S1) included all 390 viruses included as targets in the HV1 library design
528 (26), while the second only included 93 virus species (“Focal93”, Table S1), excluding any viruses
529 to which exposure within the study population was deemed to be very unlikely.

530 The *deconv* module of PepSIRF was then used with each of these linkage maps to identify
531 the most parsimonious set of virus species that can explain each set of enriched peptides. The
532 results, therefore, can be interpreted as potential seropositivities for each sample. The *deconv*
533 module accomplishes this through an iterative process. In each round, species-specific scores are
534 generated by summing the species-level scores from each enriched peptide, and the species with
535 the highest score is selected for inclusion in the output. We set a score threshold of 40; a species
536 must reach or exceed this score in order to be included in the output. Additionally, to account for
537 related species with similar scores, we allowed for ties between species if the lower scoring species
538 had a score $\geq 80\%$ of the higher scoring species (--score_tie_threshold 0.8) and if $\geq 70\%$ of the
539 enriched peptides contributing to these scores were identical between species (--
540 score_overlap_threshold 0.7). In our analysis, we considered an individual to be seropositive for
541 all tied species. Ties accounted for less than 2% of seropositivity calls. A separate *deconv* analysis
542 was run for each Z score threshold.

543

544 Identification of disparities

545 To identify significant differences in estimated seropositivity between ethnicities and/or
546 genders, we utilized a generalized linear model (GLM) implemented in Python using statsmodels
547 v3.8.8 (73). Specifically, for each viral species, we fit a binomial GLM with a single dependent
548 variable (seropositivity; 0 or 1 for each individual) and three independent variables [ethnicity
549 (categorical), gender (categorical) and age (continuous)]. We utilized an alpha of 0.05 for
550 determining significance, along with a Bonferroni correction for multiple tests (i.e., number of
551 viruses). To reduce the total number of tests, we only examined viruses with estimated population-
552 level seropositivities between 5 and 95%. These thresholds were chosen based on power
553 simulations, which indicated that, with our sample size, it would be unlikely to detect differences
554 in seropositivity $< 10\%$ as statistically significant.

555

556 Subspecies analysis of reactivity profiles

557 To assign enriched HAV peptides to individual proteins, we first aligned all 360 HAV
558 amino acid sequences from which HV1 peptides were designed using mafft v7.490 (74) with the
559 *G-INS-i method. Annotations from one of these sequences (Uniprot:Q9DWR1) were then
560 translated into alignment coordinates for the purposes of visualization (Figure 5A) and for
561 assigning peptides to proteins. A peptide was considered structural if ≥ 25 peptide amino acids
562 (83%) were assigned to HAV proteins VP1-4 and/or 2A (blue in Figure 5A). A peptide was

563 considered non-structural if ≥ 25 peptide amino acids were assigned to HAV proteins 2BC and/or
564 3ABCD (green in Figure 5A).

565
566 To determine if the disparity in EV-C seropositivity was driven by differences in
567 vaccination rate, we created a new linkage map where all peptides containing at least one 7mer
568 from any of the 2,653 poliovirus sequences (Table S3) used in the creation of the HV1 library were
569 assigned to a new taxonomic category called “poliovirus”. All EV-C peptides that did not share a
570 7mer with poliovirus sequences were assigned to an “Other EV-C” category. Next, we reran the
571 *deconv* module using this modified linkage map to estimate seropositivity for these two categories
572 separately. With this analysis, it is possible for an individual sample to be found seropositive for
573 1) just one of these categories, 2) both categories (if there are multiple enriched peptides from each
574 category), or 3) neither of these categories (even for samples previously found to be seropositive
575 for EV-C, if the enriched peptides are split between the new categories). This is a conservative
576 approach for assessing the impact of poliovirus vaccination/infection because it assumes that any
577 antibody recognizing a peptide that shares at least one amino acid 7mer with poliovirus was
578 generated in response to poliovirus infection/vaccination. In reality, these responses could have
579 been stimulated by many different EV-C strains.

580
581 To examine the contribution of different strains of EV-C to our observed “Other EV-C”
582 seropositivities, we assigned each “Other EV-C” peptide to a single subspecies group based on
583 shared amino acid 7mers, with each peptide assigned a score (between 1 and 24) equivalent to the
584 number of contained 7mers that were unique to the assigned group. To improve the sensitivity of
585 our analysis (i.e., the number of informative peptides), we focused on clades of related EV-C
586 isolates. As of December 8, 2023, the ICTV website for the Enterovirus genus
587 (<https://ictv.global/report/chapter/picornaviridae/picornaviridae/enterovirus>) listed NCBI
588 GenBank accession numbers for 27 EV-C isolates, including 6 poliovirus isolates (Table S7). We
589 downloaded polyprotein amino acid sequences for each of these and aligned them using mafft
590 v7.490 (74) with default settings. We generated a maximum-likelihood phylogeny from this
591 alignment using raxml-ng v0.5.1b (75) with the LG+FC+I+G8m model, which was selected as
592 optimal using ModelGenerator v0.85 (AIC1) (76). Based on this phylogeny, we divided the EV-C
593 subspecies into six phylogenetic clades, including five composed of non-poliovirus EV-C (EV-
594 C_1-5) (Figure 6). For each enriched “Other EV-C” peptide, we determined the number of amino
595 acid 7mers shared with the ICTV reference sequences from each EV-C clade. We assigned the
596 peptide to the clade with the highest score and normalized the associated score for this peptide by
597 subtracting the next highest clade-specific score (no assignment was made if two clades had
598 identical 7mer scores). To calculate relative reactivity scores against the five non-poliovirus clades
599 (“Relative Peptide Score” in Figure 6B, Supplemental Figure 3), each clade-specific sum of
600 enriched peptide scores was divided by the total sum of scores across all clade-assigned peptides
601 for that sample. For the ethnicity-level composite scores shown in Figure 6B (“HW” and “NHW”),
602 we summed enriched peptide scores across all HW or NHW individuals, respectively, who were

603 seropositive for our “Other EV-C” category. For these composites, individual peptide scores were
604 counted once for every sample that exhibited enrichment (i.e., a single peptide could be counted
605 multiple times if that peptide was recognized by antibodies in multiple samples).

606

607 **Acknowledgements**

608 We would like to acknowledge Sarah Namdarian for help collecting the clinical samples used in
609 this study. This work was supported by the National Institute on Minority Health and Health
610 Disparities of the National Institutes of Health under Award Number U54MD012388 and the State
611 of Arizona Technology and Research Initiative Fund (TRIF, administered by the Arizona Board
612 of Regents, through Northern Arizona University). The content is solely the responsibility of the
613 authors and does not necessarily represent the official views of the National Institutes of Health.

614

615 **Data Availability**

616 The raw peptide counts, linkage maps and data related to the figures from this study have been
617 deposited in the Open Science Framework (<https://osf.io/gvpzu/>), DOI:
618 10.17605/OSF.IO/GVPZU. All custom code is available via GitHub
619 (<https://github.com/LadnerLab>). Any additional information required to reanalyze the data
620 reported in this paper is available upon request.

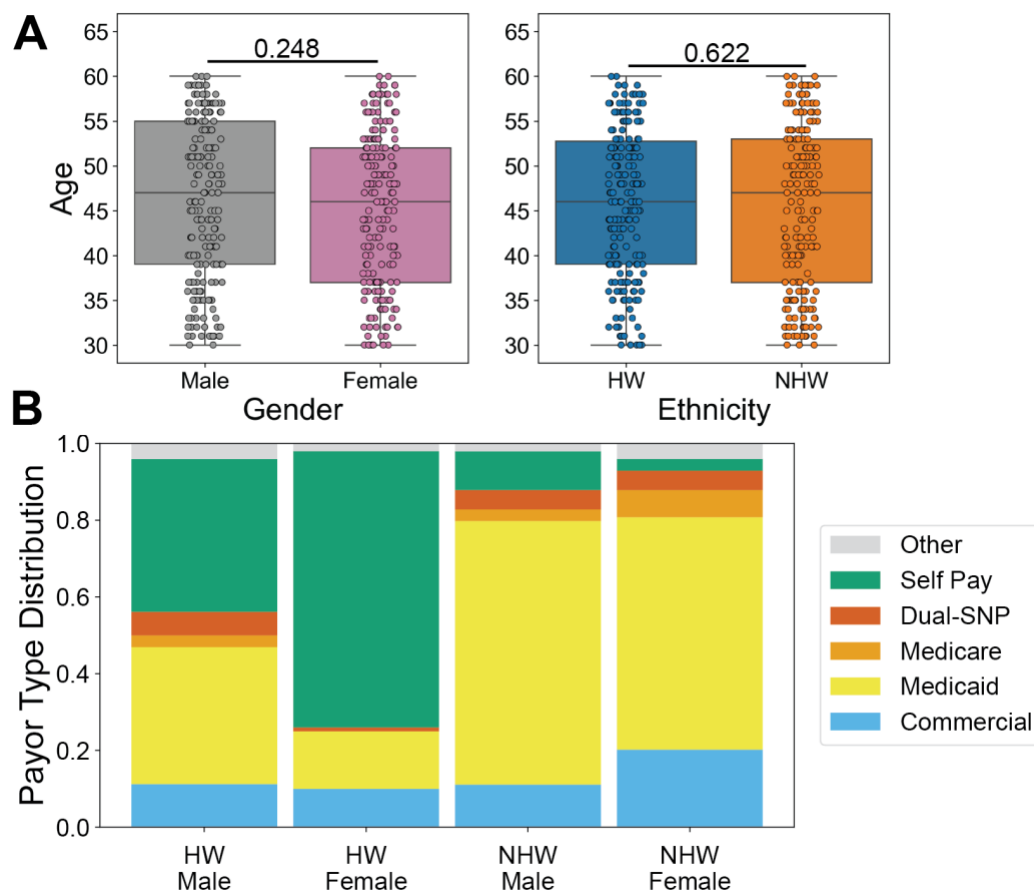
621

622

623

624 **Figures**

625



626

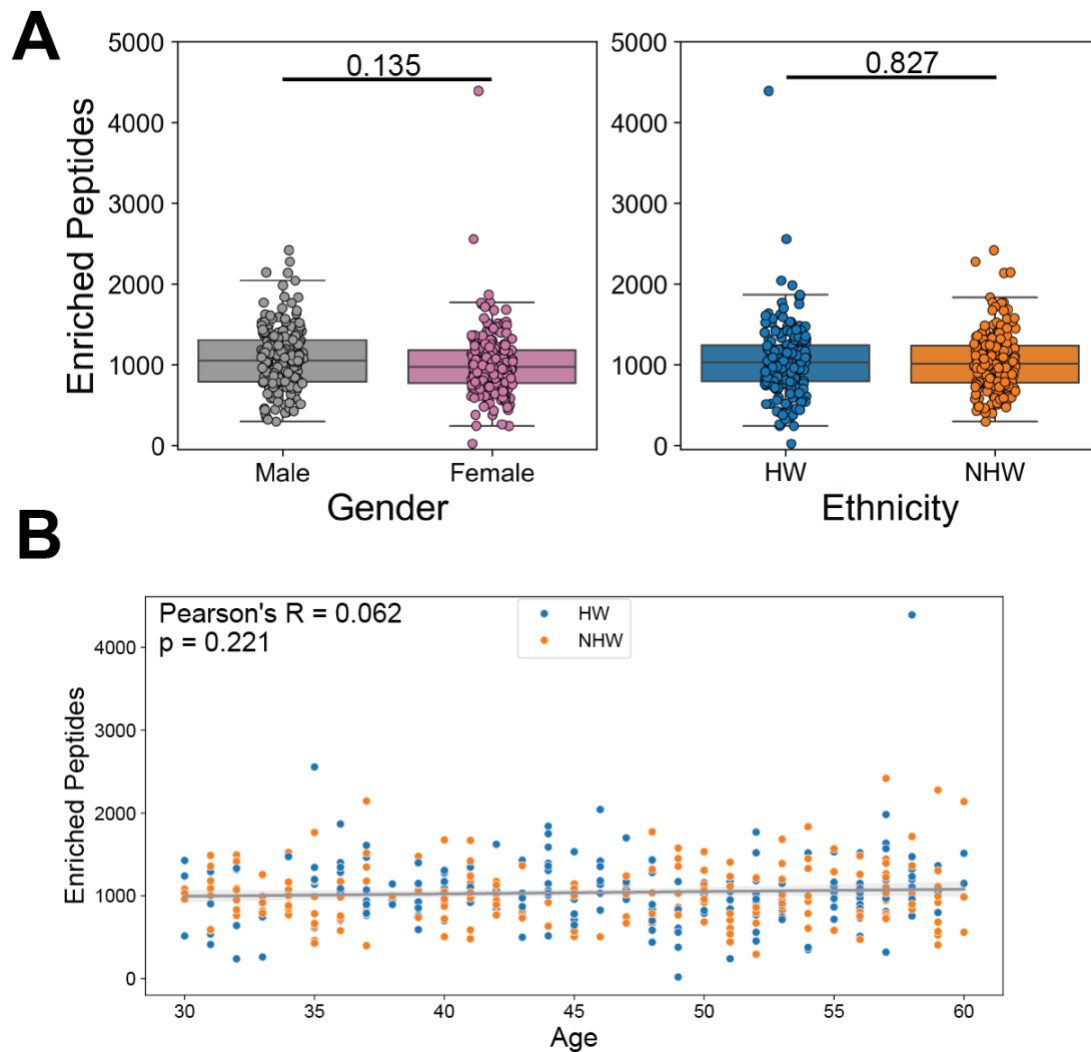
627 **Figure 1. Demographics of the study population.** Remnant serum samples were collected from
628 400 individuals at Valleywise Health in Phoenix, AZ and assayed using the HV1 PepSeq library.

629 Four samples were excluded from the analysis due to poor correlation between replicates. A) The
630 study population ranged in age from 30-60 years old and there was no statistically significant
631 difference in the age distributions between genders and ethnicities (t-tests). Individual t-test p-

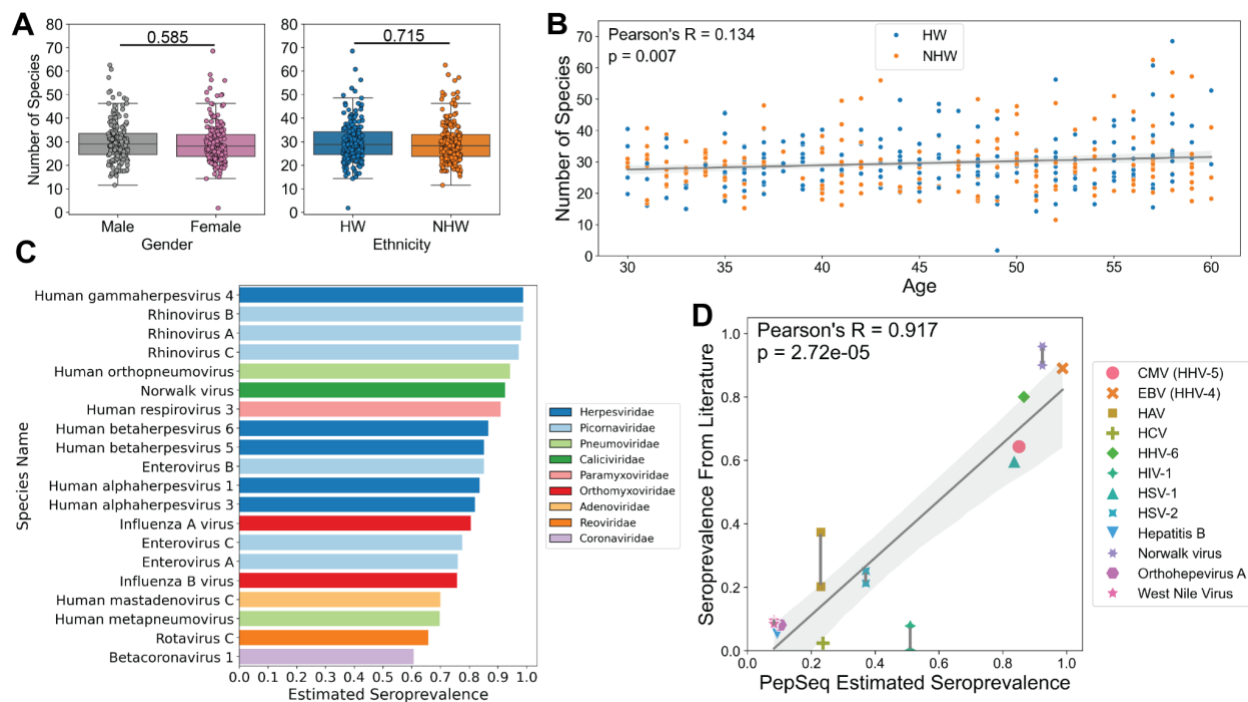
632 values comparing Male/Female and HW/NHW ages are indicated above the respective plots. Each
633 circle represents an individual. The line within each box represents the median, while the lower
634 and upper bounds of each box represent the 1st and 3rd quartiles, respectively. The whiskers extend

635 to points that lie within 1.5 interquartile ranges of the 1st and 3rd quartiles. B) Payor source varied
636 substantially among focal subpopulations. Payor source was consolidated into six general
637 categories (Table S5) and is shown for each subpopulation. Dual-SNP refers to any dual special

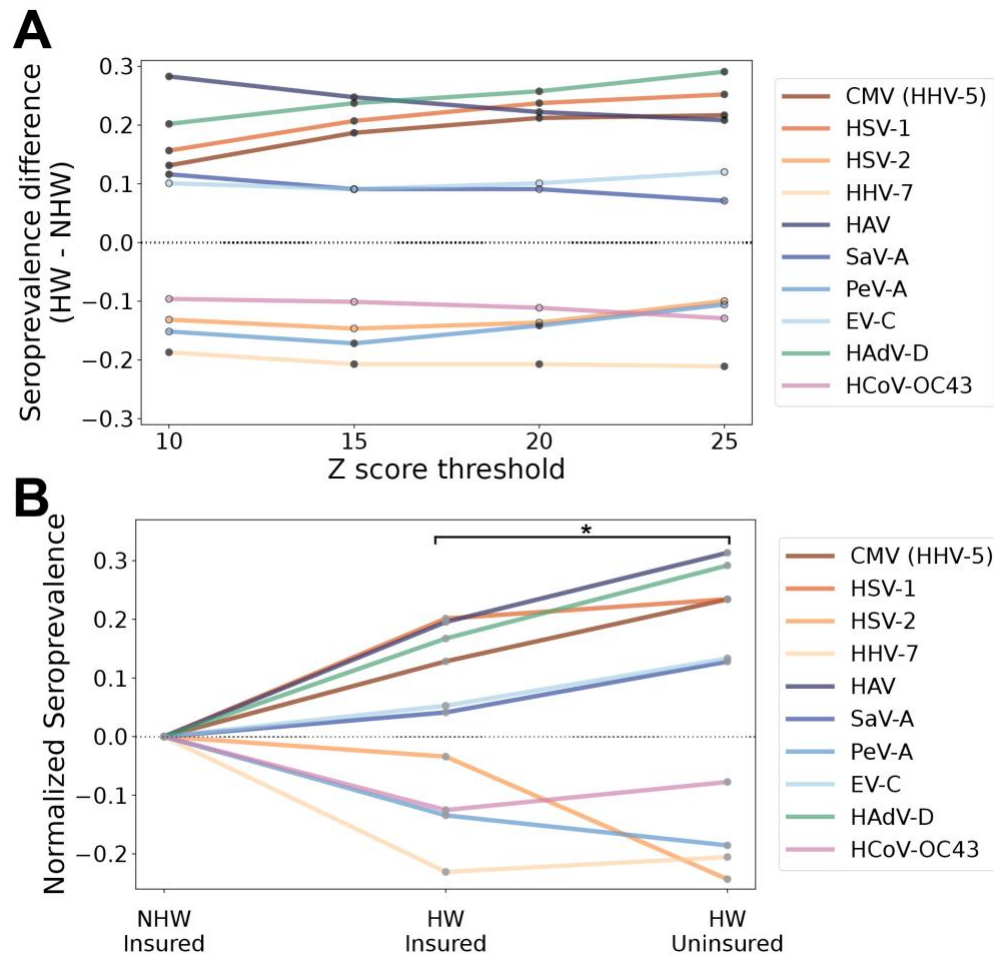
638 needs plans for individuals who qualify for both Medicare and Medicaid.



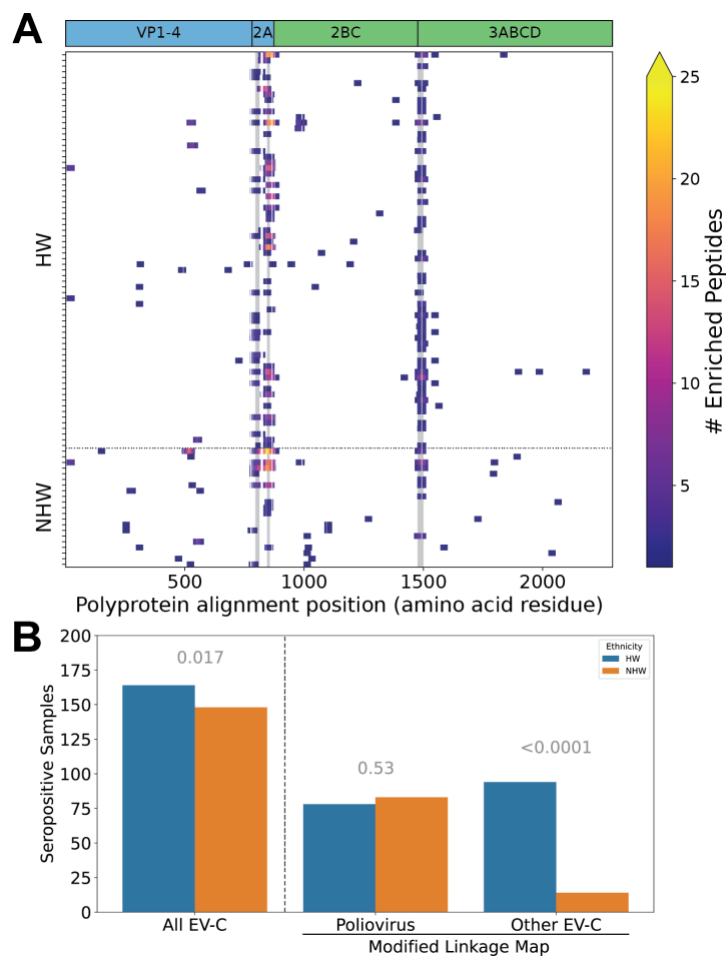
639
640 **Figure 2. PepSeq identifies similar overall levels of antibody reactivity against viral peptides**
641 **between genders and ethnicities.** A) Box plots depicting the average number of enriched PepSeq
642 peptides for each sample across the four Z score thresholds ($Z = 10, 15, 20,$ and 25). Average
643 number of enriched peptides: Female=1009.92, Male=1070.30, HW=1044.37, NHW=1035.55.
644 Individual t-test p-values comparing Male/Female and HW/NHW enriched peptide counts are
645 indicated above the respective plots. Each circle represents an individual. The line within each box
646 represents the median, while the lower and upper bounds of each box represent the 1st and 3rd
647 quartiles, respectively. The whiskers extend to points that lie within 1.5 interquartile ranges of the
648 1st and 3rd quartiles. B) Scatter plot with best-fit line showing average number of enriched
649 peptides by age. Ethnicity is indicated by the color of the points, HW=blue and NHW=orange.
650 Gray diagonal line indicates the best-fit linear regression with the shaded gray areas showing the
651 95% confidence interval. Pearson correlation was used to test for significance (p-value=0.221).
652



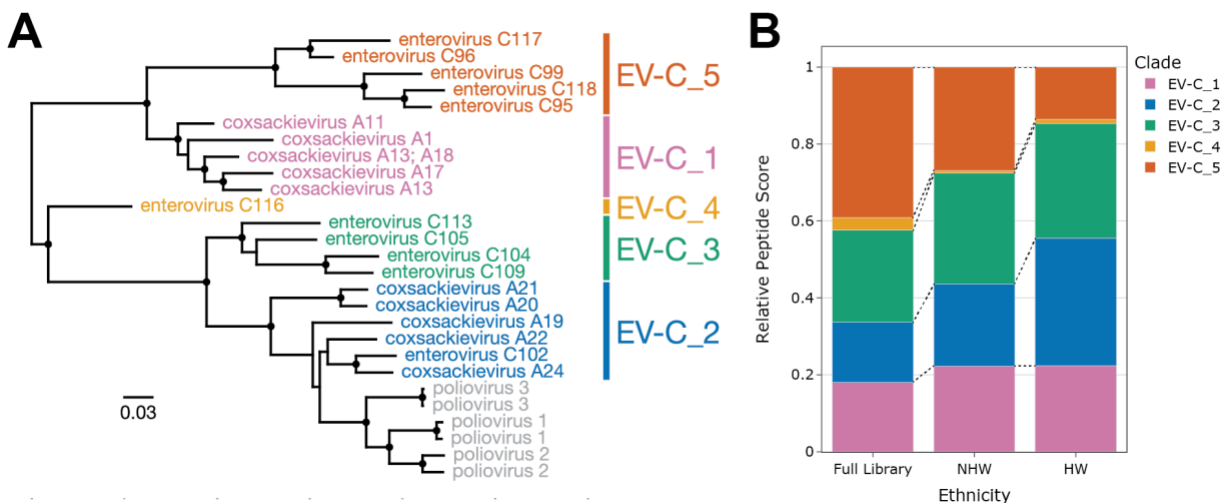
653
 654 **Figure 3. PepSeq-based estimates of seropositivity correlate with age and independent**
 655 **estimates from singleplex assays.** A) Box plots depicting the number of virus species selected as
 656 seropositive by the PepSIRF *deconv* algorithm for each sample (Focal93 linkage map), divided
 657 according to gender (left) and ethnicity (right). Average number of putatively seropositive virus
 658 species: Female=29.44, Male=29.92, HW=29.84, NHW=29.52. T-tests comparing Male/Female
 659 and HW/NHW were non-significant (p-values indicated above the respective plots). Each circle
 660 represents an individual. The line within each box represents the median, while the lower and
 661 upper bounds of each box represent the 1st and 3rd quartiles, respectively. The whiskers extend to
 662 points that lie within 1.5 interquartile ranges of the 1st and 3rd quartiles. B) Scatter plot with best-
 663 fit line showing average number of predicted seropositive virus species by age. Ethnicity is
 664 indicated by the color of the points, HW=blue and NHW=orange. Pearson correlation was used to
 665 test for significance (p-value = 0.007). Gray diagonal line indicates the best-fit linear regression
 666 with the shaded gray areas showing the 95% confidence interval. C) Estimated seroprevalence for
 667 the 20 most prevalent virus species across the entire patient cohort. Bar colors indicate the viral
 668 family. D) PepSeq estimated seroprevalence across the full dataset compared to published
 669 seroprevalence values from studies in the United States (Table S8). Different estimates for the
 670 same virus are connected by a vertical line. The gray diagonal line indicates the best-fit linear
 671 regression with the shaded gray areas showing the 95% confidence interval. Pearson's R value and
 672 p-value shown in the top left. Abbreviations: RSV=respiratory syncytial virus, EBV=Epstein-Barr
 673 virus, HHV=Human herpesvirus HSV=herpes simplex virus, CMV=cytomegalovirus,
 674 HAV=hepatitis A virus, HIV=human immunodeficiency virus, HCV=hepatitis C virus.



675
 676 **Figure 4. Significant differences in seroprevalence by ethnicity and payor status.** A) Line plot
 677 depicting ten species with a significant difference (p -value<0.05) in seropositivity between HW
 678 and NHW calculated by fitting a generalized linear model at each Z score threshold before
 679 (outlined points) and after (filled points) Bonferroni correction for multiple tests. Negative
 680 differences indicate higher seroprevalence in NHWs and positive differences indicate higher
 681 seroprevalence in HWs. B) Line plot depicting normalized seroprevalence for the same ten viruses
 682 shown in (A), with values calculated separately for insured and uninsured individuals.
 683 Seroprevalence is being shown for a Z score threshold of 15 and was normalized against the value
 684 for insured NHWs. Asterisk indicates the significant increase of the absolute value of the
 685 normalized seroprevalence for uninsured HWs compared to insured HWs across all 10 viruses
 686 (paired t-test p -value=0.013). Abbreviations: CMV=cytomegalovirus, HSV=herpes simplex virus,
 687 HHV=human herpesvirus, HAV=hepatitis A virus, SaV-A = salivirus A, PeV-A=parechovirus A,
 688 EV-C=*Enterovirus C*, HAdV-D=human adenovirus D, HCoV-OC43=human coronavirus OC43.
 689



690
 691 **Figure 5. Ethnicity-based disparities in viral infection rates are not driven by vaccination**
 692 **against hepatitis A virus (HAV) or poliovirus.** A) Enriched peptides and public epitopes
 693 identified in both structural and non-structural HAV proteins. Heatmap of enriched peptides across
 694 the HAV proteome for all samples seropositive for HAV with a Z score threshold of 15. Samples
 695 are broken up by ethnicity on the y-axis and position within the HAV polyprotein is shown on the
 696 x-axis (amino acid residues, alignment coordinates). The three most commonly reactive epitopes
 697 within these samples are highlighted with gray vertical markings. The positions of individual viral
 698 proteins are indicated across the top; blue=structural (vaccination or infection) and green=non-
 699 structural (infection only). B) Bar plot showing the number of seropositive HW (total n=198) and
 700 NHW (total n=198) individuals. Seropositivity was calculated using the original linkage map
 701 (Focal93) with poliovirus peptides assigned to the Enterovirus C (EV-C) virus species (“All EV-
 702 C”), or with a modified linkage map where peptides sharing ≥ 1 7mer with polioviruses were
 703 removed from the EV-C species and treated as their own category (“Poliovirus” and “Other EV-
 704 C”). P-values for binomial GLM comparing the impact of ethnicity (HW, NHW) on seropositivity
 705 are indicated in gray above each viral category.
 706



707
 708 **Figure 6. Antibody reactivity against clade-specific EV-C peptides.** A) Maximum-likelihood
 709 phylogenetic tree of all 27 ICTV listed EV-C isolates based on an amino acid alignment of the full
 710 polyprotein. Branch lengths indicate relative levels of amino acid divergence with the scale bar
 711 indicating the equivalent of 0.03 changes per site. Black circles indicate nodes with bootstrap
 712 support ≥ 80 . Six phylogenetic clades are identified by the different colors. Genbank accession
 713 numbers for each sequence can be found in Table S7. B) Bar plot showing clade-specific relative
 714 peptide scores for all "Other EV-C" peptides present in the HV1 PepSeq library (Full Library, null
 715 distribution) compared to the subset of enriched peptides for all HW and NHW individuals
 716 predicted to be seropositive against "Other EV-C" (Figure 5B). Each color represents one of the
 717 "Other EV-C" phylogenetic clades shown in (A).
 718

719

720 **References**

- 721 1. Tyler KL. 2018. Acute Viral Encephalitis. *N Engl J Med* 379:557–566.
- 722 2. Naniche D, Oldstone MB. 2000. Generalized immunosuppression: how viruses undermine
723 the immune response. *Cell Mol Life Sci* 57:1399–1407.
- 724 3. Pereira L. 2018. Congenital Viral Infection: Traversing the Uterine-Placental Interface.
725 *Annu Rev Virol* 5:273–299.
- 726 4. Lu S, Huang X, Liu R, Lan Y, Lei Y, Zeng F, Tang X, He H. 2022. Comparison of COVID-
727 19 Induced Respiratory Failure and Typical ARDS: Similarities and Differences. *Front Med*
728 9:829771.
- 729 5. Shahrizaila N, Lehmann HC, Kuwabara S. 2021. Guillain-Barré syndrome. *Lancet*
730 397:1214–1228.
- 731 6. Feldstein LR, Rose EB, Horwitz SM, Collins JP, Newhams MM, Son MBF, Newburger
732 JW, Kleinman LC, Heidemann SM, Martin AA, Singh AR, Li S, Tarquinio KM, Jaggi P,
733 Oster ME, Zackai SP, Gillen J, Ratner AJ, Walsh RF, Fitzgerald JC, Keenaghan MA,
734 Alharash H, Doymaz S, Clouser KN, Giuliano JS Jr, Gupta A, Parker RM, Maddux AB,
735 Havalad V, Ramsingh S, Bukulmez H, Bradford TT, Smith LS, Tenforde MW, Carroll CL,
736 Riggs BJ, Gertz SJ, Daube A, Lansell A, Coronado Munoz A, Hobbs CV, Marohn KL,
737 Halasa NB, Patel MM, Randolph AG, Overcoming COVID-19 Investigators, CDC COVID-
738 19 Response Team. 2020. Multisystem Inflammatory Syndrome in U.S. Children and
739 Adolescents. *N Engl J Med* 383:334–346.
- 740 7. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, Pujol JC, Klaser
741 K, Antonelli M, Canas LS, Molteni E, Modat M, Jorge Cardoso M, May A, Ganesh S,
742 Davies R, Nguyen LH, Drew DA, Astley CM, Joshi AD, Merino J, Tsereteli N, Fall T,
743 Gomez MF, Duncan EL, Menni C, Williams FMK, Franks PW, Chan AT, Wolf J, Ourselin
744 S, Spector T, Steves CJ. 2021. Attributes and predictors of long COVID. *Nat Med* 27:626–
745 631.
- 746 8. Tobin NH, Campbell AJP, Zerr DM, Melvin AJ. 2011. Chapter 95 - Life-Threatening Viral
747 Diseases and Their Treatment, p. 1324–1335. *In* Fuhrman, BP, Zimmerman, JJ (eds.),
748 *Pediatric Critical Care (Fourth Edition)*. Mosby, Saint Louis.
- 749 9. Filippi CM, von Herrath MG. 2008. Viral trigger for type 1 diabetes: pros and cons.
750 *Diabetes* 57:2863–2871.
- 751 10. Brown JJ, Jabri B, Dermody TS. 2018. A viral trigger for celiac disease. *PLoS Pathog*
752 14:e1007181.
- 753 11. Mitra AK, Clarke K. 2010. Viral obesity: fact or fiction? *Obesity Reviews*
754 <https://doi.org/10.1111/j.1467-789x.2009.00677.x>.

- 755 12. Bar-Or A, Pender MP, Khanna R, Steinman L, Hartung H-P, Maniar T, Croze E, Aftab BT,
756 Giovannoni G, Joshi MA. 2020. Epstein-Barr Virus in Multiple Sclerosis: Theory and
757 Emerging Immunotherapies. *Trends Mol Med* 26:296–310.
- 758 13. Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, Elledge SJ, Niebuhr DW,
759 Scher AI, Munger KL, Ascherio A. 2022. Longitudinal analysis reveals high prevalence of
760 Epstein-Barr virus associated with multiple sclerosis. *Science* 375:296–301.
- 761 14. Sarid R, Gao S-J. 2011. Viruses and human cancer: From detection to causality. *Cancer*
762 *Letters* <https://doi.org/10.1016/j.canlet.2010.09.011>.
- 763 15. Devanand DP. 2018. Viral Hypothesis and Antiviral Treatment in Alzheimer’s Disease.
764 *Curr Neurol Neurosci Rep* 18:55.
- 765 16. Liu G, Markowitz LE, Hariri S, Panicker G, Unger ER. 2015. Seroprevalence of 9 Human
766 Papillomavirus Types in the United States, 2005–2006. *J Infect Dis* 213:191–198.
- 767 17. Xu F, Sternberg MR, Kottiri BJ, McQuillan GM, Lee FK, Nahmias AJ, Berman SM,
768 Markowitz LE. 2006. Trends in herpes simplex virus type 1 and type 2 seroprevalence in
769 the United States. *JAMA* 296:964–973.
- 770 18. CDC. 2021. HIV and Gay and Bisexual Men.
771 <https://www.cdc.gov/hiv/group/msm/index.html>. Retrieved 15 August 2022.
- 772 19. Purcell DW, Johnson CH, Lansky A, Prejean J, Stein R, Denning P, Gau Z, Weinstock H,
773 Su J, Crepaz N. 2012. Estimating the population size of men who have sex with men in the
774 United States to obtain HIV and syphilis rates. *Open AIDS J* 6:98–107.
- 775 20. Erhart LM, Ernst KC. 2012. The changing epidemiology of hepatitis A in Arizona
776 following intensive immunization programs (1988-2007). *Vaccine* 30:6103–6110.
- 777 21. World Health Organization. 2009. Dengue: Guidelines for Diagnosis, Treatment,
778 Prevention and Control. World Health Organization.
- 779 22. 2022. Statistics and maps - 2020. <https://www.cdc.gov/dengue/statistics-maps/2020.html>.
780 Retrieved 23 August 2022.
- 781 23. Hans R, Marwaha N. 2014. Nucleic acid testing-benefits and constraints. *Asian J Transfus*
782 *Sci* 8:2–3.
- 783 24. Lequin RM. 2005. Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay
784 (ELISA). *Clin Chem* 51:2415–2418.
- 785 25. Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung’u T, Ruxrungtham K, Sanchez J, Brander
786 C, Chung RT, O’Connor KC, Walker B, Larman HB, Elledge SJ. 2015. Viral immunology.
787 Comprehensive serological profiling of human populations using a synthetic human virome.
788 *Science* 348:aaa0698.

- 789 26. Ladner JT, Henson SN, Boyle AS, Engelbrektsen AL, Fink ZW, Rahee F, D’ambrozio J,
790 Schaecher KE, Stone M, Dong W, Dadwal S, Yu J, Caligiuri MA, Cieplak P, Bjørås M,
791 Fenstad MH, Nordbø SA, Kainov DE, Muranaka N, Chee MS, Shiryaev SA, Altin JA.
792 2021. Epitope-resolved profiling of the SARS-CoV-2 antibody response identifies cross-
793 reactivity with endemic human coronaviruses. *Cell Rep Med* 2:100189.
- 794 27. Elko EA, Nelson GA, Mead HL, Kelley EJ, Carvalho ST, Sarbo NG, Harms CE, Le Verche
795 V, Cardoso AA, Ely JL, Boyle AS, Piña A, Henson SN, Rahee F, Keim PS, Celona KR, Yi
796 J, Settles EW, Bota DA, Yu GC, Morris SR, Zaia JA, Ladner JT, Altin JA. 2022. COVID-
797 19 vaccination elicits an evolving, cross-reactive antibody response to epitopes conserved
798 with endemic coronavirus spike proteins. *Cell Rep* 40:111022.
- 799 28. Kelley EJ, Henson SN, Rahee F, Boyle AS, Engelbrektsen AL, Nelson GA, Mead HL,
800 Anderson NL, Razavi M, Yip R, Ladner JT, Scriba TJ, Altin JA. 2023. Virome-wide
801 detection of natural infection events and the associated antibody dynamics using
802 longitudinal highly-multiplexed serology. *Nat Commun* 14:1783.
- 803 29. Henson SN, Elko EA, Swiderski PM, Liang Y, Engelbrektsen AL, Piña A, Boyle AS, Fink
804 Z, Facista SJ, Martinez V, Rahee F, Brown A, Kelley EJ, Nelson GA, Raspet I, Mead HL,
805 Altin JA, Ladner JT. 2023. PepSeq: a fully in vitro platform for highly multiplexed serology
806 using customizable DNA-barcoded peptide libraries. *Nat Protoc* 18:396–423.
- 807 30. Mohan D, Wansley DL, Sie BM, Noon MS, Baer AN, Laserson U, Larman HB. 2018.
808 PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes.
809 *Nat Protoc* 13:1958–1978.
- 810 31. Patricia O’Campo JB. 2004. *Eliminating Health Disparities: Measurement and Data Needs*.
811 National Academies Press.
- 812 32. Staras SAS, Dollard SC, Radford KW, Flanders WD, Pass RF, Cannon MJ. 2006.
813 Seroprevalence of cytomegalovirus infection in the United States, 1988-1994. *Clin Infect*
814 *Dis* 43:1143–1151.
- 815 33. Balfour HH Jr, Sifakis F, Sliman JA, Knight JA, Schmeling DO, Thomas W. 2013. Age-
816 specific prevalence of Epstein-Barr virus infection among individuals aged 6-19 years in the
817 United States and factors affecting its acquisition. *J Infect Dis* 208:1286–1293.
- 818 34. Klevens RM, Kruszon-Moran D, Wasley A, Gallagher K, McQuillan GM, Kuhnert W,
819 Teshale EH, Drobeniuc J, Bell BP. 2011. Seroprevalence of hepatitis A virus antibodies in
820 the U.S.: results from the National Health and Nutrition Examination Survey. *Public Health*
821 *Rep* 126:522–532.
- 822 35. McQuillan GM, Kruszon-Moran D, Kottiri BJ, Curtin LR, Lucas JW, Kington RS. 2004.
823 Racial and ethnic differences in the seroprevalence of 6 infectious diseases in the United
824 States: data from NHANES III, 1988-1994. *Am J Public Health* 94:1952–1958.
- 825 36. Okuno T, Takahashi K, Balachandra K, Shiraki K, Yamanishi K, Takahashi M, Baba K.
826 1989. Seroepidemiology of human herpesvirus 6 infection in normal children and adults. *J*

- 827 Clin Microbiol 27:651–653.
- 828 37. St Louis ME, Rauch KJ, Petersen LR, Anderson JE, Schable CA, Dondero TJ. 1990.
829 Seroprevalence rates of human immunodeficiency virus infection at sentinel hospitals in the
830 United States. The Sentinel Hospital Surveillance Group. *N Engl J Med* 323:213–218.
- 831 38. Kirby AE, Kienast Y, Zhu W, Barton J, Anderson E, Sizemore M, Vinje J, Moe CL. 2020.
832 Norovirus Seroprevalence among Adults in the United States: Analysis of NHANES Serum
833 Specimens from 1999-2000 and 2003-2004. *Viruses* 12.
- 834 39. Cangin C, Focht B, Harris R, Strunk JA. 2019. Hepatitis E seroprevalence in the United
835 States: Results for immunoglobulins IGG and IGM. *J Med Virol* 91:124–131.
- 836 40. Cahill ME, Yao Y, Nock D, Armstrong PM, Andreadis TG, Diuk-Wasser MA,
837 Montgomery RR. 2017. West Nile Virus Seroprevalence, Connecticut, USA, 2000-2014.
838 *Emerg Infect Dis* 23:708–710.
- 839 41. Schweitzer BK, Kramer WL, Sambol AR, Meza JL, Hinrichs SH, Iwen PC. 2006.
840 Geographic factors contributing to a high seroprevalence of West Nile virus-specific
841 antibodies in humans following an epidemic. *Clin Vaccine Immunol* 13:314–318.
- 842 42. Ye C, Luo J, Wang X, Xi J, Pan Y, Chen J, Yang X, Li G, Sun Q, Yang J. 2017.
843 Development of a peptide ELISA to discriminate vaccine-induced immunity from natural
844 infection of hepatitis A virus in a phase IV study. *Eur J Clin Microbiol Infect Dis* 36:2165–
845 2170.
- 846 43. Alfaro-Murillo JA, Ávila-Agüero ML, Fitzpatrick MC, Crystal CJ, Falleiros-Arlant L-H,
847 Galvani AP. 2020. The case for replacing live oral polio vaccine with inactivated vaccine in
848 the Americas. *Lancet* 395:1163–1166.
- 849 44. Kang X, Li Y, Fan L, Lin F, Wei J, Zhu X, Hu Y, Li J, Chang G, Zhu Q, Liu H, Yang Y.
850 2012. Development of an ELISA-array for simultaneous detection of five encephalitis
851 viruses. *Virol J* 9:56.
- 852 45. Fukushi S. 2020. Competitive ELISA for the Detection of Serum Antibodies Specific for
853 Middle East Respiratory Syndrome Coronavirus (MERS-CoV). *Methods Mol Biol*
854 2203:55–65.
- 855 46. Boncristiani HF, Criado MF, Arruda E. 2009. Respiratory Viruses, p. 500–518. *In*
856 *Encyclopedia of Microbiology*. Elsevier.
- 857 47. Dunn SR, Ryder AB, Tollefson SJ, Xu M, Saville BR, Williams JV. 2013.
858 Seroepidemiologies of human metapneumovirus and respiratory syncytial virus in young
859 children, determined with a new recombinant fusion protein enzyme-linked immunosorbent
860 assay. *Clin Vaccine Immunol* 20:1654–1656.
- 861 48. Parrott RH, Vargosko AJ, Kimhw, Bell JA, Chanock RM. 1962. Acute respiratory diseases
862 of viral etiology. III. parainfluenza. Myxoviruses. *Am J Public Health Nations Health*

- 863 52:907–917.
- 864 49. DeGroot NP, Haynes AK, Taylor C, Killerby ME, Dahl RM, Mustaqim D, Gerber SI,
865 Watson JT. 2020. Human parainfluenza virus circulation, United States, 2011–2019.
866 *Journal of Clinical Virology* <https://doi.org/10.1016/j.jcv.2020.104261>.
- 867 50. Mc Quillan G, Kruszon-Moran D, Flagg EW, Paulose-Ram R. 2016. Prevalence of Herpes
868 Simplex Virus Type 1 and Type 2 in Persons Aged 14–49: United States, 2015–2016.
- 869 51. Rubicz R, Leach CT, Kraig E, Dhurandhar NV, Grubbs B, Blangero J, Yolken R, Göring
870 HH. 2011. Seroprevalence of 13 common pathogens in a rapidly growing U.S. minority
871 population: Mexican Americans from San Antonio, TX. *BMC Res Notes* 4:433.
- 872 52. Ernst KC, Erhart LM. 2014. The role of ethnicity and travel on Hepatitis A vaccination
873 coverage and disease incidence in Arizona at the United States-Mexico Border. *Hum
874 Vaccin Immunother* 10:1396–1403.
- 875 53. Lazcano-Ponce E, Conde-Gonzalez C, Rojas R, DeAntonio R, Romano-Mazzotti L,
876 Cervantes Y, Ortega-Barria E. 2013. Seroprevalence of hepatitis A virus in a cross-sectional
877 study in Mexico: Implications for hepatitis A vaccination. *Hum Vaccin Immunother* 9:375.
- 878 54. Li L, Victoria J, Kapoor A, Blinkova O, Wang C, Babrzadeh F, Mason CJ, Pandey P, Triki
879 H, Bahri O, Oderinde BS, Baba MM, Bukbuk DN, Besser JM, Bartkus JM, Delwart EL.
880 2009. A novel picornavirus associated with gastroenteritis. *J Virol* 83:12002–12006.
- 881 55. Kitajima M, Iker BC, Rachmadi AT, Haramoto E, Gerba CP. 2014. Quantification and
882 genetic analysis of salivirus/klassevirus in wastewater in Arizona, USA. *Food Environ Virol*
883 6:213–216.
- 884 56. Pauly M, Hoppe E, Mugisha L, Petrzalkova K, Akoua-Koffi C, Couacy-Hymann E, Anoh
885 AE, Mossoun A, Schubert G, Wiersma L, Pascale S, Muyembe J-J, Karhemere S, Weiss S,
886 Leendertz SA, Calvignac-Spencer S, Leendertz FH, Ehlers B. 2014. High prevalence and
887 diversity of species D adenoviruses (HAdV-D) in human populations of four Sub-Saharan
888 countries. *Virol J* 11:25.
- 889 57. Karelehto E, Brouwer L, Benschop K, Kok J, Basile K, McMullan B, Rawlinson W, Druce
890 J, Nicholson S, Selvarangan R, Harrison C, Lankachandra K, van Eijk H, Koen G, de Jong
891 M, Pajkrt D, Wolthers KC. 2019. Seroepidemiology of Parechovirus A3 Neutralizing
892 Antibodies, Australia, the Netherlands, and United States. *Emerg Infect Dis* 25:148–152.
- 893 58. Sridhar A, Karelehto E, Brouwer L, Pajkrt D, Wolthers KC. 2019. Parechovirus A
894 Pathogenesis and the Enigma of Genotype A-3. *Viruses* 11.
- 895 59. Zaki SR, Keating MK. 2018. 13 - Viral Diseases, p. 244–288. *In* Zander, DS, Farver, CF
896 (eds.), *Pulmonary Pathology (Second Edition)*. Elsevier, Philadelphia.
- 897 60. Oliver S, James SH. 2017. Herpesviruses, p. 621–627. *In* *International Encyclopedia of*
898 *Public Health*. Elsevier.

- 899 61. Epstein LG, Shinnar S, Hesdorffer DC, Nordli DR, Hamidullah A, Benn EKT, Pellock JM,
900 Frank LM, Lewis DV, Moshe SL, Shinnar RC, Sun S, FEBSTAT study team. 2012. Human
901 herpesvirus 6 and 7 in febrile status epilepticus: the FEBSTAT study. *Epilepsia* 53:1481–
902 1488.
- 903 62. Li Y, Qu T, Li D, Jing J, Deng Q, Wan X. 2022. Human herpesvirus 7 encephalitis in an
904 immunocompetent adult and a literature review. *Virology* 19:200.
- 905 63. 2020. Immigrants in Arizona. American Immigration Council.
- 906 64. Wald A, Corey L. Persistence in the population: epidemiology, transmission, p. . *In* Arvin,
907 A, Campadelli-Fiume, G, Mocarski, E, Moore, PS, Roizman, B, Whitley, R, Yamanishi, K
908 (eds.), *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*. Cambridge
909 University Press, Cambridge.
- 910 65. Cao G, Jing W, Liu J, Liu M. 2021. The global trends and regional differences in incidence
911 and mortality of hepatitis A from 1990 to 2019 and implications for its prevention. *Hepatology*
912 *Int* 15:1068–1082.
- 913 66. Zuhair M, Smit GSA, Wallis G, Jabbar F, Smith C, Devleeschauwer B, Griffiths P. 2019.
914 Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and
915 meta-analysis. *Rev Med Virol* 29:e2034.
- 916 67. Robertson BH, Jia XY, Tian H, Margolis HS, Summers DF, Ehrenfeld E. 1992. Serological
917 approaches to distinguish immune response to hepatitis A vaccine and natural infection.
918 *Vaccine* 10 Suppl 1:S106–9.
- 919 68. Tokarz R, Haq S, Sameroff S, Howie SRC, Lipkin WI. 2013. Genomic analysis of
920 coxsackieviruses A1, A19, A22, enteroviruses 113 and 104: viruses representing two clades
921 with distinct tropism within enterovirus C. *J Gen Virol* 94:1995–2004.
- 922 69. Brouwer L, Benschop KSM, Nguyen D, Kamau E, Pajkrt D, Simmonds P, Wolthers KC.
923 2020. Recombination Analysis of Non-Poliiovirus Members of the Enterovirus C Species;
924 Restriction of Recombination Events to Members of the Same 3DPol Cluster. *Viruses* 12.
- 925 70. Fink ZW, Martinez V, Altin J, Ladner JT. 2020. PepSIRF: a flexible and comprehensive
926 tool for the analysis of data from highly-multiplexed DNA-barcoded peptide assays. *arXiv*
927 preprint arXiv:2007.05050.
- 928 71. Brown AM, Bolyen E, Raspet I, Altin JA, Ladner JT. 2022. PepSIRF + QIIME 2: software
929 tools for automated, reproducible analysis of highly-multiplexed serology data
930 <https://doi.org/10.48550/ARXIV.2207.11509>.
- 931 72. Mina MJ, Kula T, Leng Y, Li M, de Vries RD, Knip M, Siljander H, Rewers M, Choy DF,
932 Wilson MS, Larman HB, Nelson AN, Griffin DE, de Swart RL, Elledge SJ. 2019. Measles
933 virus infection diminishes preexisting antibodies that offer protection from other pathogens.
934 *Science* 366:599–606.

- 935 73. Seabold S, Perktold J. 2010. Statsmodels: Econometric and statistical modeling with
936 pythonProceedings of the 9th Python in Science Conference. SciPy.
- 937 74. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
938 improvements in performance and usability. *Mol Biol Evol* 30:772–780.
- 939 75. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast,
940 scalable and user-friendly tool for maximum likelihood phylogenetic inference.
941 *Bioinformatics* 35:4453–4455.
- 942 76. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO. 2006. Assessment of
943 methods for amino acid matrix selection and their use on empirical data shows that ad hoc
944 assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29.
- 945