

The Journal of Infectious Diseases

Viral genetic determinants of prolonged respiratory syncytial virus infection among infants in a healthy term birth cohort.

--Manuscript Draft--

Manuscript Number:	
Full Title:	Viral genetic determinants of prolonged respiratory syncytial virus infection among infants in a healthy term birth cohort.
Short Title:	RSV variants and prolonged infection
Article Type:	Major Article
Section/Category:	Viruses
Keywords:	RSV, variants, prolonged, infection, GWAS
Corresponding Author:	Dylan Lawless, PhD École Polytechnique Fédérale de Lausanne Lausanne, VD SWITZERLAND
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	École Polytechnique Fédérale de Lausanne
Corresponding Author's Secondary Institution:	
First Author:	Dylan Lawless, PhD
First Author Secondary Information:	
Order of Authors:	Dylan Lawless, PhD Christopher G. McKennan, PhD Suman R. Das, PhD Thomas Junier, PhD Zhi Ming Xu, MPhil Larry J Anderson, MD Tebeb Gebretsadik, MPH Meghan H. Shilts, MHS, MS Emma Larkin, PhD Christian Rosas-Salazar, MD, MPH James D. Chappell, MD, PhD Jacques Fellay, MD, PhD Tina V. Hartert, MD, MPH
Order of Authors Secondary Information:	
Manuscript Region of Origin:	UNITED STATES
Abstract:	<p>Background: Respiratory syncytial virus (RSV) is associated with acute respiratory infection. We sought to identify RSV variants associated with prolonged infection.</p> <p>Methods: Among healthy term infants we identified those with prolonged RSV infection and conducted 1) a viral GWAS for association with prolonged RSV infection using RSV whole-genome sequencing, 2) a human GWAS to test the dependence of infection risk on genotype, 3) analysis of all viral public sequences, 4) assessment of immunological responses, and 5) a summary of all major functional data. Analyses were adjusted for viral/human population structure and host factors associated with</p>

	<p>infection risk.</p> <p>Results: We identified p.E123K/D and p.P218T/S/L in G protein that were associated with prolonged infection ($P_{adj} = 0.01$). We found no evidence of host genetic risk for infection. The RSV variant positions approximate sequences that could bind a putative viral receptor, heparan sulfate.</p> <p>Conclusions: Using analysis of both viral and host genetics we identified a novel RSV variant associated with prolonged infection in healthy infants and no evidence supporting host genetic susceptibility to infection. As the capacity of RSV for chronicity and its viral reservoir are not defined, these findings are important for understanding the impact of RSV on chronic disease and endemicity.</p>
Suggested Reviewers:	<p>Mitch Grayson Professor, Nationwide Children's Mitchell.Grayson@Nationwidechildrens.org Subject expert</p>
	<p>Steve Varga Professor, University of Iowa steven-varga@uiowa.edu Subject expert</p>
	<p>Edward Walsh Professor, University of Rochester edward.walsh@rochesterregional.org Subject expert</p>
	<p>John Devincenzo Dr, UT Memphis jdevince@uthsc.edu Subject expert</p>
	<p>Louis Bont Professor, Utrecht lbont@umcutrecht.nl Subject expert</p>
	<p>Ralph Tripp Professor, Georgia ratripp@uga.edu Subject expert</p>

July 29, 2022

RE: Original research submission, "Viral genetic determinants of prolonged respiratory syncytial virus infection among infants in a healthy term birth cohort"

Dear Editor,

We are submitting our manuscript "Viral genetic determinants of prolonged respiratory syncytial virus infection among infants in a healthy term birth cohort".

In this manuscript we tackle an observation which is of great interest in our field but not yet reported, determining if there are respiratory syncytial virus (RSV) variants associated with prolonged infection. We conducted biweekly RSV surveillance in nearly 2,000 healthy term infants and identified those with prolonged RSV infection. We applied both host and pathogen genetic association testing to detect variants that contribute to prolonged infection. This consisted of [1] a viral GWAS, [2] a human GWAS, [3] analysis of all public sequence data, and [4] assessing local immunological responses. We identified viral genetic variants associated with prolonged infection which we assessed in depth.

Our results are novel and represent an in-depth comprehensive computational statistical analysis of both host and viral genetics providing compelling evidence for RSV viral variants associated with RSV persistence in healthy human infants, a finding of significant importance to understanding the impact of RSV on chronic disease development and viral endemicity.

Thank you for your consideration of this manuscript for *The Journal of Infectious Diseases*.

All authors have seen and approved the manuscript, contributed significantly to the work, and the manuscript has neither been previously published nor is being considered for publication elsewhere.

Sincerely yours,

Dylan Lawless, PhD, dylan.lawless@epfl.ch,
Tina V. Hartert, MD, MPH tina.hartert@vumc.org

1 Viral genetic determinants of prolonged respiratory syncytial virus infection
2 among infants in a healthy term birth cohort.

3 Dylan Lawless, PhD¹, Christopher G. McKennan, PhD², Suman R. Das, PhD₃, Thomas Junier, PhD₄, Zhi
4 Ming Xu, MPhil¹, Larry J Anderson, MD⁵, Tebeb Gebretsadik, MPH⁶, Meghan H. Shilts, MHS, MS⁷, Emma
5 Larkin, PhD⁸, Christian Rosas-Salazar, MD, MPH₈, James D. Chappell, MD, PhD⁹, Jacques Fellay, MD,
6 PhD^{1,10}, and Tina V. Hartert, MD, MPH^{7,9}

7

8 ¹Global Health Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne,
9 Lausanne, Switzerland, ²Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania,
10 United States of America, ³Division of Infectious Diseases, Department of Medicine, Vanderbilt University
11 Medical Center, Nashville, Tennessee, United States of America, ⁴Swiss Institute of Bioinformatics, Vital-
12 IT Group, Switzerland, ⁵Department of Pediatrics, Emory University School of Medicine and Children's
13 Healthcare of Atlanta, Atlanta, Georgia, United States of America, ⁶Department of Biostatistics, Vanderbilt
14 University Medical Center, Nashville, Tennessee, United States of America, ⁷Department of Medicine,
15 Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, ⁸Division of
16 Allergy, Immunology, and Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt
17 University Medical Center, Nashville, Tennessee, United States of America, ⁹Department of Pediatrics,
18 Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, ¹⁰Biomedical Data
19 Science Center, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland.¹

20

21 **Running title**

22 RSV variants and prolonged infection

23

24 **Key points**

25 Using a comprehensive computational analysis of viral and host genetics we identified a novel RSV
26 variant associated with prolonged infection and no evidence supporting host genetic infection
27 susceptibility, findings important to understanding RSV contribution to chronic disease and viral
28 endemicity.

29

30 **Abbreviations**

¹ All authors have completed a conflict-of-interest form (COI). There were no COI. Funding sources are listed in the funding sources section.

31 ALT (alternative); CI (confidence interval); GWAS (genome-wide association study); G (glycoprotein); H
32 (hemagglutinin); HN (hemagglutinin-neuraminidase); IFN (interferon); IQR (interquartile range); INSPIRE
33 (The INFant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure); LD (linkage
34 disequilibrium); LRTI (lower respiratory tract infection); MAF (minor allele frequency); MFI (median
35 fluorescence intensity); MSA (multiple sequence alignment); OR (odds ratio); PCR (polymerase chain
36 reaction); PCA (Principal component analysis); REF (reference); RT (reverse transcription); SVD (singular
37 value decomposition); SNP (single nucleotide polymorphism); VE (variance explained); MSA (multiple
38 sequence alignment); RSV (respiratory syncytial virus).

39

40 Notice of Prior Presentation

41 The results of the host genome wide association study analyses included in this manuscript were
42 presented during the European Society of Human Genetics Conference in June 2020 in Berlin, Germany,
43 which was held remotely [1].

44

45 Ethics Statement for Human Subjects Research

46 The protocol and informed consent documents were approved by the Institutional Review Board at
47 Vanderbilt University Medical Center (#111299). One parent of each participant in the cohort study
48 provided written informed consent for participation in this study. The informed consent document
49 explained study procedures and use of data and biospecimens for future studies, including genetic
50 studies.

51

52 Competing interests

53 All authors have completed a conflict-of-interest form (COI). Summary of any COI: There were no COI.

54

55 Funding sources

56 Funding was supplied from the US National Institutes of Health and Swiss National Science Foundation.
57 Grants: U19 AI 095227, UG3/UH3 OD023282, UL1 TR002243, SNSF IZSEZ0_191968 (TVH), SNSF
58 310030L_197721 (JF), X01 HLG244 RS&G (EL).

59

60 **Summary**

61 A comprehensive computational statistical analysis of both host and viral genetics provided compelling
62 evidence for RSV viral persistence in healthy human infants, a finding of significant importance to
63 understanding the impact of RSV on chronic disease and viral endemicity.

64

65 **1 Abstract**

66 Background: Respiratory syncytial virus (RSV) is associated with acute respiratory infection. We sought to
67 identify RSV variants associated with prolonged infection.

68

69 Methods: Among healthy term infants we identified those with prolonged RSV infection and conducted 1)
70 a viral GWAS for association with prolonged RSV infection using RSV whole-genome sequencing, 2) a
71 human GWAS to test the dependence of infection risk on genotype, 3) analysis of all viral public
72 sequences, 4) assessment of immunological responses, and 5) a summary of all major functional data.
73 Analyses were adjusted for viral/human population structure and host factors associated with infection
74 risk.

75

76 Results: We identified p.E123K/D and p.P218T/S/L in G protein that were associated with prolonged
77 infection ($P_{adj} = 0.01$). We found no evidence of host genetic risk for infection. The RSV variant positions
78 approximate sequences that could bind a putative viral receptor, heparan sulfate.

79

80 Conclusions: Using analysis of both viral and host genetics we identified a novel RSV variant associated
81 with prolonged infection in healthy infants and no evidence supporting host genetic susceptibility to
82 infection. As the capacity of RSV for chronicity and its viral reservoir are not defined, these findings are
83 important for understanding the impact of RSV on chronic disease and endemicity.

84

85 **2 Introduction**

86 Human orthopneumovirus, formerly known (and still referred to) as respiratory syncytial virus (RSV),
87 results in significant global morbidity and mortality [2]. By the age of two to three years, nearly all children
88 have been infected with RSV at least once [3]. RSV is a seasonal mucosal pathogen that primarily infects
89 upper and lower respiratory tract epithelium, although it has been recovered from non-airway sources [4–
90 9]. While RSV is mainly associated with acute respiratory infection, many RNA viruses can establish
91 prolonged or persistent infection in some infected individuals [10]. Prolonged shedding of RSV, especially
92 in young infants and following first infection, has been demonstrated, with longer average duration of viral
93 shedding when polymerase chain reaction (PCR) is used to detect RSV [11]. While younger age and first
94 infection are associated with protracted infection [3; 12], it is not known whether specific viral factors
95 contribute to prolonged RSV infection in infants. This is important, as prolonged infection may contribute
96 to enhanced transmission and developmental changes to the early life airway epithelium. Further, the

97 reservoir of RSV infection is not understood, and it is possible that some RSV strains sustain a low level
98 of ongoing viral circulation in the community until seasonal or other influences favor epidemic spread [13].
99

100 The objectives of this study were therefore to determine if there exist host genetic risk alleles for RSV
101 infection and to identify viral genetic variation associated with prolonged infection. These motivating
102 questions are of fundamental interest in understanding viral and host genetic contributions that may
103 underlie the development of chronic respiratory morbidity due to RSV, including asthma.

104

105 **3 Methods**

106 The protocol and informed consent documents were approved by the Institutional Review Board at
107 Vanderbilt University Medical Center (#111299). One parent of each participant in the cohort study
108 provided written informed consent for participation in this study. The informed consent document
109 explained study procedures and use of data and biospecimens for future studies, including genetic
110 studies.

111

112 Among healthy term infants in a cohort specifically designed to capture first RSV infection we identified
113 those with prolonged RSV infection and conducted 1) a viral GWAS using RSV whole-genome
114 sequencing to determine the relationship between viral genotypes and prolonged infant RSV infection, 2)
115 a human GWAS to test the dependence of first year RSV infection risk on the genotype, 3) an analysis of
116 all viral public sequence data, 4) an assessment of the local immunological RSV responses, and 5) a
117 summary of all the major functional data for the identified viral variant. Full details of the methods are
118 included in the Supplement, sections [8.1 - 8.13](#).

119

120 **4 Results**

121 **4.1 Cohort characteristics**

122 The INSPIRE cohort consisted of 1,949 enrolled infants among whom there were 2,093 in-person
123 respiratory illness visits completed during winter virus season, November – March, of each year (Figure
124 [S1](#)); the median (interquartile range [IQR]) number of in-person respiratory illness visits per infant during
125 this surveillance window was 1 [[1](#); [2](#)]. There were 344 RSV PCR-positive samples from 325 individuals
126 which were sequenced. Prolonged infection was *a priori* defined as repeatedly meeting criteria for acute
127 respiratory infection with RSV PCR positive nasal samples \geq 15 days between testing. There were 19
128 infants who met the definition of prolonged infection with available viral sequencing used to confirm
129 clonality of original and subsequent virus detections. The mean RSV CT value of first infections was 25.9 ± 7.1 ,
130 and second detection was 31.6 ± 5.4 . The mean number of days between detections was 25 ± 25
131 days (Figure [S2](#)). Table [1](#) lists the cohort characteristics of infants with prolonged RSV infection
132 compared with other RSV infection and the entire cohort.

133

134 4.2 Host genetic analyses

135 We explored whether RSV infection in infancy is a natural assignment (quasi-random) event and, unlike
136 severity of early-life RSV infection [14], occurs independently of host genetics. For the candidate SNP
137 analysis, we considered childhood asthma- and RSV LRTI-associated SNPs identified in Pividori et al.
138 [15]; Janssen et al. [16]; Pasanen et al. [17]. The first is the largest childhood asthma GWAS to date, and,
139 to our knowledge, the latter 2 represent the most comprehensive studies of RSV LRTI-associated SNPs.
140 To further reduce the multiple testing burden, we only analysed SNPs with MAF ≥ 0.1 in at least one of
141 the White, Black, or Hispanic ethnicity groups. Associations between genotype at the resulting 54 SNPs
142 (50 childhood asthma- and 4 RSV LRTI-associated SNPs) and RSV infection in infancy in our data are
143 given in Figure 1. The data are consistent with little to no effect of genotype at these SNPs on RSV
144 infection in infancy.

145

146 We further investigated the possibility that the analysis was underpowered to identify associations with
147 these SNPs by pooling information across SNPs to estimate the average genetic effect size [18]. We
148 estimated the narrow-sense heritability of RSV infection during infancy on the latent liability scale (h^2),
149 which, if > 0 , would indicate an accumulation of small genetic effects. We estimated h^2 to be exactly 0,
150 suggesting that, if present, infant RSV infection-related genetic signals are both small and sparse
151 (Supplemental sec 8.6).

152

153 4.3 Population structure

154 A summary of protein coding genes in RSV is illustrated in Figure 2 A. Our analysis focused on F and G
155 protein. The phylogenetic tree based on multiple sequence alignment (MSA) of G protein amino acid
156 sequences is shown in Figure 2 B. One obvious feature causing a separation in genetic diversity is G
157 protein partial gene duplication, which has emerged in recent years within RSV-A strains [19]. RSV-B
158 strains with a homologous duplication have existed for two decades, although the selection process
159 leading to emergence and clinical implications have not been entirely defined.

160

161 PCA was used for reducing the dimensionality of sequence data, where PC1 accounted for 95.19% of
162 cumulative variance, and variance attributed to other PCs was roughly uniformly distributed (Figure 2 C).
163 We observed prolonged infections by viruses from different phylogenetic clades, rather than one specific
164 clade (Figure 2 C), indicating that these results are not confounded by latent clade membership.

165

166 4.4 Genetic invariance of prolonged infection

167 The duration of RSV shedding in Kenyan infants has been reported previously [13]. Based on these
168 findings, infection events separated by at least 15 days with symptoms were expected to be “new”
169 infections [13]. Figure 2 D panel [i] summarizes every pairwise genetic distance between every viral

170 sequence, where small distances indicate pairs with closely related sequences. Panels [ii] and [iii], which
171 summarize the difference in sequence similarity distributions between viruses from the same host and
172 different hosts, show that RSV sequences corresponding to initial and subsequent viral detections are
173 nearly identical. These results support the conclusion that such cases are prolonged (i.e., failure to clear)
174 infections rather than new infections.

175

176 4.5 Variants in G glycoprotein significantly associated with prolonged infection

177 The consensus sequence within the cohort was assigned based on the major allele. Variants at the amino
178 acid level were defined as either reference (REF) or alternative (ALT) and assessed for their association
179 with prolonged infection. The model consisted of (A) the binary response (prolonged infection Yes/No),
180 and (B) predictors; (1) viral genotype (REF/ALT amino acid), (2) viral PCs 1-5, (3) host sex, and host
181 features that have been previously demonstrated as significantly associated with infection; (4) self-
182 reported race/ethnicity, (5) child-care attendance, or living with another child \leq 6 years of age at home
183 [20]. A significant genetic association was identified between prolonged infection and the lead variant
184 after Bonferroni correction for multiple testing (threshold for number independent variants $< 0.05/23 =$
185 0.002), as shown in Figure 3 A, p value = 0.0006.

186

187 To determine whether this association was simply due to population stratification between strains A and
188 B, a subset analysis was performed using independently assessed clinical laboratory strain labels for A
189 and B. The same direction of effect indicated that the association was not a false positive, although in this
190 significantly smaller sub-analysis the result was not significant.

191

192 To assess the possibility of a false positive association due to population structure within our cohort, we
193 assessed the magnitude of variance explained (VE) at every amino acid position. Figure 3 B (panel [i])
194 shows the variance explained by each amino acid in PCs1-5. The cumulative proportion of variance for
195 PCs 1-5 was 99.5% (PC1 = 95%, PC2 = 3%). The values are illustrated according to protein position in
196 panels [ii-iii]. The lead association variant had 0.603% VE for PC1 and 0.458% VE for PC2, a negligible
197 effect that precludes spurious association by allele frequency between populations.

198

199 After identifying a significant viral genetic association with prolonged infection, we quantified the
200 correlation of variants with the lead proxy. Clumping was performed with ranking based on MAF and with
201 a cut-off threshold of $r^2 \geq 0.8$ (Supplemental Figure S3). The association model was repeated for all
202 variants, defining protein p.E123K/D and p.P218T/S/L as candidate causal variants associated with
203 prolonged infection as shown in Figure 3 C. No other variants were correlated with this outcome.

204

205 To determine whether p.E123K/D and p.P218T/S/L variant genotypes are novel and potentially influence
206 viral fitness, we searched the public viral data repository of NCBI Human orthopneumovirus, taxid:11250,

207 which contained data from 27 countries worldwide, sample collection dates from 1956 onward, and 1084
208 glycoprotein protein sequences after curation. The variants were present at a low and stable frequency,
209 without obvious temporal enrichment (Supplemental Figure S4). Thus, while historical data reveal no
210 positive selective advantage attached to p.E123K/D and p.P218T/S/L, longstanding circulation and
211 linkage in prolonged RSV infection suggest that these polymorphisms are present in the viral inoculum
212 and do not arise through recurrent mutational events.

213

214 Due to multiple testing correction according to our analysis plan, an association also originally identified in
215 F protein was rejected and therefore omitted from further discussion. For posterity, the variant position
216 was p.N116S (relative to strain A GenBank: AMN91253.1).

217

218 4.6 Functional interpretation

219 Cell-attachment proteins of paramyxoviruses (G protein in RSV) span the viral envelope and form spike-
220 like projections from the virion surface. RSV G protein is a type II integral membrane protein consisting of
221 298 amino acid residues comprising N-terminal cytoplasmic (p.1-43), transmembrane helical (p.43-63),
222 and extracellular (p.64-298) domains (Figure 3 D). RSV G protein ectodomain also exists in a soluble
223 secreted form, p.66 – 298, which functions in immune evasion [21–23]. G protein interacts with the small
224 hydrophobic (SH) protein [24] and, via the N-terminus, with matrix (M) [25] protein. It has also been
225 reported to form homo-oligomers [26]. The variant amino acid positions associated with prolonged
226 infection reside in a portion of the G protein ectodomain of unassigned specific function and linearly non-
227 contiguous with sequences that bind cell-surface heparan sulfate, which likely promotes RSV cell-
228 attachment (p.187-198) [21–23]. In addition, these positions do not contribute to known neutralization
229 epitopes on G protein. Information available in PDB was insufficient to infer effects of p.E123K/D and
230 p.P218T/S/L on local or regional protein structure. The potential effect on glycosylation is indeterminate.
231 Figure 3 D illustrates the position of these variants relative to summarised known functional features.

232

233 4.7 Host response

234 Prolonged infections associated with G protein variants p.E123K/D and p.P218T/S/L were on average
235 less severe compared with other circulating variants, and all were limited to the upper respiratory tract
236 (Table 1). Therefore, we analysed nasal wash samples collected during acute RSV infection for a panel of
237 cytokines involved in antiviral immune responses and observed differential IFN α and IFN γ levels
238 segregating according to viral antigenic group—A or B. Both cytokines were elevated in group B
239 infections compared to group A. The groups A and B median (lower-and upper-quartile) values were 9.5
240 (3-22.5) and 12.6 (4.1-25.8) median fluorescence intensity (MFI), respectively, for IFN α and 3.6 (1-7) and
241 4 (2-7.4) MFI, respectively, for IFN γ (group A, n = 149; group B, n = 103). As prolonged infections with
242 p.E123K/D and p.P218T/S/L genotypes were exclusively group B, the dichotomous relationship of IFN α

243 and IFNy levels to antigenic group precluded evaluation of G protein variants as independent predictors
244 of IFN α and IFNy production.

245

246 5 Discussion

247 In this study of term healthy infants, we found no evidence of host genetic susceptibility to RSV infection
248 during infancy. This allowed our analysis to focus on elucidation of viral drivers of prolonged infection. A
249 significant viral genetic association in the RSV G protein, p.E123K/D and p.P218T/S/L, with prolonged
250 infant RSV infection was identified. These variants were not associated with severe disease, and public
251 data reveal their consistent presence at low frequencies over the past 30 years, without evidence of
252 enrichment by positive selective pressure over time. The two variants we identified in G are correlated
253 with non-random association analogous to LD in the human diploid genome and therefore not likely
254 random mutations, but instead co-inherited in the infecting inoculum. This suggests an evolutionary
255 benefit and raises the question of why such variants have maintained a stable but low frequency in the
256 human population for at least four decades. These strains are a potential reservoir, emerging seasonally
257 in response to immune, environmental, or other forces. Alternatively, the polymorphisms might recurrently
258 arise de novo during infection of some individuals but are poorly transmissible because of suboptimal
259 fitness. The possibility of viral mutational immune escape has been reported for infants who struggle to
260 control primary RSV infections, allowing for prolonged viral replication and not previously described viral
261 rebound [27].

262

263 The RSV variants associated with prolonged infection in our cohort, G p.E123K/D and p.P218T/S/L, lie in
264 the extracellular region, and there are no known mechanistic features that directly overlap, although it is
265 possible that variant positions approximate sequences that bind a putative viral receptor, heparan sulfate
266 [22], in the G protein three-dimensional structure. G protein amino acid positions 123 and 218 are not part
267 of known antibody neutralization epitopes or CD8+ cytotoxic T-cell epitopes (Figure 3 D). In addition to
268 heparan sulfate, interactions between viral G protein and CX3CR1, the receptor for the CX3C chemokine
269 fractalkine, have been reported to modulate the immune response and facilitate infection [21– 23; 28–30].
270 Furthermore, the mature secreted isoform of G protein (p.66-298) is thought to facilitate viral antibody
271 evasion by acting as an antigen decoy and modifying the activity of leukocytes bearing Fc-gamma
272 receptors [31]. Our findings raise the interesting prospect that G protein variants associated with
273 prolonged infection alter a key interaction at the immune interface between pathogen and host.

274

275 Although this study was not designed to define mechanisms underlying the association of G protein
276 variants with prolonged infection, these sequence changes might dampen antiviral immune responses
277 and thereby delay viral clearance. Although we observed differences in the acute antiviral response
278 between subjects with resolved and prolonged infection, specifically increased levels of types 1 and 2 IFN
279 in nasal secretions, we could not make causal inference about variant sequences because of

280 confounding by co-linearity of these polymorphisms with RSV antigenic group. Results of nasal cytokine
281 analysis are nevertheless consistent with a contemplated role for altered immune responses in extended
282 infections by G protein variant strains [32]. It is also possible that strains harbouring G protein p.E123K/D
283 and p.P218T/S/L variants are cleared more slowly and foster an immune environment of low-level chronic
284 stimulation or exhaustion. We previously demonstrated that infants infected with RSV in their first year of
285 life have dampened subsequent antiviral immune responses in early childhood [33] as well as changes in
286 airway epithelial cell metabolism [34].

287

288 While this study has a number of significant strengths, including one of few population-based surveillance
289 studies of first RSV infections during infancy among term healthy infants, our findings are also subject to
290 some limitations. First, this study was not designed with the primary intention to examine infection
291 duration, and additional sampling following initial RSV infection was triggered by a repeat acute
292 respiratory illness. Asymptomatic prolonged infections would therefore not have been captured. Second,
293 our study cohort was small, necessitating focus on viral surface glycoproteins, F and G, due to their
294 variability and importance in host immunity. A larger cohort with serial sampling would be required to
295 diminish the impact of co-linearity of viral genotypes with antigenic groups and to perform informative viral
296 whole genome analysis. Genome-wide information might elucidate other determinants of prolonged
297 infection or pathogen fitness that mediate and/or modulate effects of phenotype-driving variations. Third,
298 again due to small sample size, we could only investigate host genetic risk for infection, not prolonged
299 infection. While we have not specifically assessed subjects for rare monogenic variants that may underlie
300 immunodeficiency, our enrolment criteria included only infants who were term and otherwise healthy.
301 While we performed an interaction analysis for the outcome of host asthma, host genetics, and pathogen
302 genetics and found no significant interaction, our sample size is unlikely sufficient to exclude such an
303 interaction. Lastly, while we do not expect a role for immune memory in these first-in-life RSV infections,
304 we cannot exclude modulatory effects of maternal antibody, which we did not measure. Despite these
305 limitations, the results are novel and represent an in-depth comprehensive computational statistical
306 analysis of both host and viral genetics providing compelling evidence for RSV viral strain persistence in
307 healthy human infants, a finding of significant importance to understanding the impact of RSV on chronic
308 disease and viral endemicity.

309

310 In summary, we identified a novel RSV viral variant associated with prolonged infection in healthy infants,
311 but no evidence of host genetic susceptibility to infant RSV infection. Understanding host and viral
312 mechanisms that contribute to prolonged infection will be important in crafting strategies to control the
313 short- and long-term impact of RSV infection. The identification of RSV variants associated with
314 prolonged infection might also improve vaccine design, particularly if these variants stimulate robust
315 immunity or, in contrast, escape the immune response or induce immunopathologic conditions. The
316 growing availability of large genomic and functional data sources provides opportunities for advancing our

317 understanding of the pathogenesis of infant RSV infection, defining the contribution of viral genetic
318 variants to acute and chronic disease, and informing the development of effective vaccines. As neither
319 the capacity of RSV for prolonged infection in immunocompetent hosts nor a viral reservoir has been
320 delineated, these results are of fundamental interest in understanding viral and host genetic contributions
321 that may promote prolonged infection and influence development of chronic respiratory morbidity.

322

323 **6 Links**

324 **6.1 Software**

325 R v4.1.0 was used for data preparation and analysis <http://www.r-project.org>.
326 R package caret was used for analysis: genetic correlations.
327 R package dplyr was used for data curation.
328 R package factoextra was used for analysis: PCA, and to visualise eigenvalues and variance. R package
329 ggplot2 was used for data visualisation.
330 R package MASS was used to analysis: logistic regression model data.
331 R package stats was used for analysis: including glm for logistic regressions. R package stringr was used
332 for data curation.
333 R package tidyr was used for data curation.
334 asn2fsa <https://www.huge-man-linux.net/man1/asn2fsa.html>
335 clc novo assemble qiagenbioinformatics.com
336 Clustal Omega <https://www.ebi.ac.uk/Tools/msa/clustalo/>
337 dbNSFP (database) <http://database.liulab.science/dbNSFP> [35] GCTA
338 <https://cnsgenomics.com/software/gcta/> [36]
339 GenBank <https://www.ncbi.nlm.nih.gov/genbank/>
340 IQ-Tree <https://www.iqtree.org/> [37]
341 KING <https://people.virginia.edu/~wc9c/KING/> [38]
342 MAFFT <https://mafft.cbrc.jp/alignment/software/> [39]
343 NextAlign <https://github.com/nextstrain/nextclade>
344 PLINK <http://zzz.bwh.harvard.edu/plink/> [40]
345 Tbl2asn <https://www.ncbi.nlm.nih.gov/genbank/tbl2asn2/>
346 Viral Genome ORF Reader, VIGOR 3.0 <https://sourceforge.net/projects/jcvi-vigor/files/>
347 RCSB PDB <https://www.rcsb.org> UniProt <https://www.uniprot.org>
348

349 **6.2 Data sources**

350 Dataset <https://www.ncbi.nlm.nih.gov/bioproject/267583>.
351 Dataset <https://www.ncbi.nlm.nih.gov/bioproject/225816>.
352 J. Craig Venter Institute <https://www.jcvi.org>.

353 GenBank:NC 001989 Bovine orthopneumovirus, complete genome
354 https://www.ncbi.nlm.nih.gov/nuccore/NC_001989.
355 Reference data <https://www.ncbi.nlm.nih.gov/gene/?term=1489824>. G attachment glycoprotein [Human
356 orthopneumovirus]; ID: 1489824; Location: NC 001781.1 (4675..5600); Aliases: HRSVgp07.
357 Reference data <https://www.ncbi.nlm.nih.gov/gene/?term=37607642>. G attachment glycoprotein [Human
358 orthopneumovirus]; ID: 37607642; Location: NC 038235.1 (4673..5595); Aliases: DZD21 gp07.
359 Reference data for all public NCBI Virus <https://www.ncbi.nlm.nih.gov/labs/virus/vssi/> for species: Human
360 orthopneumovirus; genus: Orthopneumovirus; family: Pneumoviridae.
361 Reference data
362 https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Human%20orthopneumovirus,%20taxid:11250 contains sequence data for Virus Lineage ss=Human
364 orthopneumovirus, taxid:11250 nucleotide: 26'965, protein: 53'804, RefSeq Genomes: 2.
365 Reference https://www.ncbi.nlm.nih.gov/protein/NP_056862.1 GCF 002815475.1 (release 2018-08-19)
366 Nucleotide Accessions: NC 038235.1, protein: Y 009518856.1
367 Reference https://www.ncbi.nlm.nih.gov/protein/YP_009518856.1 GCF 000855545.1 (release 2015-02-
368 12) Nucleotide Accessions: NC 001781.1, protein: NP 056862.1 (strain B1).
369

370 7 Code availability

371 Public upload of analysis code on GitHub <https://github.com/DylanLawless/inspire2022lawless.github.io>.
372

373 References

- 374 [1] D Lawless, C Rosas-Salazar, T Gebretsadik, K Turi, B Snyder, P Wu, J Fellay, and T Hartert.
375 Genome-wide association study of susceptibility to respiratory syncytial virus infection during infancy. In
376 European Journal of Human Genetics, volume 28, pages 319–319. Springer Nature Campus, 4 Crinan
377 St, London, N1 9XW, England, 2020.
- 378 [2] C. B. Hall, G. A. Weinberg, M. K. Iwane, A. K. Blumkin, K. M. Edwards, M. A. Staat, P. Auinger, M. R.
379 Griffin, K. A. Poehling, D. Erdman, C. G. Grijalva, Y. Zhu, and P. Szilagyi. The burden of respiratory
380 syncytial virus infection in young children. N Engl J Med, 360(6):588–98, February 2009. ISSN 0028-4793
381 (Print) 0028-4793. doi: 10.1056/NEJMoa0804877. Edition: 2009/02/07.
- 382 [3] W Paul Glezen, Larry H Taber, Arthur L Frank, and Julius A Kasel. Risk of primary infection and
383 reinfection with respiratory syncytial virus. American journal of diseases of children, 140(6):543–546,
384 1986.
- 385 [4] V. Bokun, J. J. Moore, R. Moore, C. C. Smallcombe, T. J. Harford, F. Rezaee, F. Esper, and G.
386 Piedimonte. Respiratory syncytial virus exhibits differential tropism for distinct hu- man placental cell types
387 with Hofbauer cells acting as a permissive reservoir for infection. PLoS One, 14(12):e0225767, 2019.
388 ISSN 1932-6203. doi: 10.1371/journal.pone.0225767. Edition: 2019/12/04.
- 389 [5] H. A. Cubie, L. A. Duncan, L. A. Marshall, and N. M. Smith. Detection of respiratory syncytial virus
390 nucleic acid in archival postmortem tissue from infants. Pediatr Pathol Lab Med, 17(6):927–38, November
391 1997. ISSN 1077-1042 (Print) 1077-1042. Edition: 1997/11/14.
- 392 [6] D. Nadal, W. Wunderli, O. Meurmann, J. Briner, and J. Hirsig. Isolation of respiratory syncytial virus
393 from liver tissue and extrahepatic biliary atresia material. Scand J Infect Dis, 22(1):91–3, 1990. ISSN
394 0036-5548 (Print) 0036-5548. doi: 10.3109/00365549009023125. Edition: 1990/01/01.

- 395 [7] D. R. O'Donnell, M. J. McGarvey, J. M. Tully, I. M. Balfour-Lynn, and P. J. Openshaw. Respiratory
396 syncytial virus RNA in cells from the peripheral blood during acute infection. *J Pediatr*, 133(2):272–4,
397 August 1998. ISSN 0022-3476 (Print) 0022-3476. doi: 10.1016/s0022-3476(98)70234-3. Edition:
398 1998/08/26.
- 399 [8] F. Rezaee, L. F. Gibson, D. Piktel, S. Othumpangat, and G. Piedimonte. Respiratory syncytial virus
400 infection in human bone marrow stromal cells. *Am J Respir Cell Mol Biol*, 45(2):277–86, August 2011.
401 ISSN 1044-1549 (Print) 1044-1549. doi: 10.1165/rcmb.2010-0121OC. Edition: 2010/10/26.
- 402 [9] A. Rohwedder, O. Kaminer, J. Forster, K. Schneider, E. Schneider, and H. Werchau. Detection of
403 respiratory syncytial virus RNA in blood of neonates by polymerase chain reaction. *J Med Virol*,
404 54(4):320–7, April 1998. ISSN 0146-6615 (Print) 0146-6615. doi: 10.1002/(sici)1096-
405 9071(199804)54:4<320::aid-jmv13>3.0.co;2-j. Edition: 1998/04/29.
- 406 [10] Richard E Randall and Diane E Griffin. Within host rna virus persistence: mechanisms and
407 consequences. *Current opinion in virology*, 23:35–42, 2017.
- 408 [11] P. K. Munywoki, D. C. Koech, C. N. Agoti, N. Kibirige, J. Kipkoech, P. A. Cane, G. F. Medley, and D.
409 J. Nokes. Influence of age, severity of infection, and co-infection on the duration of respiratory syncytial
410 virus (RSV) shedding. *Epidemiol Infect*, 143(4):804–12, March 2015. ISSN 0950-2688 (Print) 0950-2688.
411 doi: 10.1017/s0950268814001393. Edition: 2014/06/06.
- 412 [12] Bindya Bagga, L Harrison, P Roddam, and JP DeVincenzo. Unrecognized prolonged viral replication
413 in the pathogenesis of human rsv infection. *Journal of Clinical Virology*, 106: 1–6, 2018.
- 414 [13] Emelda A Okiro, Lisa J White, Mwanajuma Ngama, Patricia A Cane, Graham F Medley, and D
415 James Nokes. Duration of shedding of respiratory syncytial virus in a community study of kenyan children.
416 *BMC infectious diseases*, 10(1):1–7, 2010.
- 417 [14] Emma K Larkin and Tina V Hartert. Genes associated with rsv lower respiratory tract infection and
418 asthma: the application of genetic epidemiological methods to understand causality. *Future virology*,
419 10(7):883–897, 2015.
- 420 [15] Milton Pividori, Nathan Schoettler, Dan L Nicolae, Carole Ober, and Hae Kyung Im. Shared and
421 distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-
422 wide studies. *The Lancet Respiratory Medicine*, 7(6): 509–522, 2019.
- 423 [16] Riny Janssen, Louis Bont, Christine LE Siezen, Hennie M Hodemaekers, Marieke J Erm- ers, Gerda
424 Doornbos, Ruben van't Slot, Ciska Wijmenga, Jelle J Goeman, Jan LL Kim- pen, et al. Genetic
425 susceptibility to respiratory syncytial virus bronchiolitis is predominantly associated with innate immune
426 genes. *Journal of Infectious Diseases*, 196(6): 826–834, 2007.
- 427 [17] Anu Pasanen, Minna K Karjalainen, Louis Bont, Eija Pippo-Savolainen, Marja Ruot- salainen, Emma
428 Goks ör, Kuldeep Kumawat, Hennie Hodemaekers, Kirsi Nuolivirta, Tuomas Jartti, et al. Genome-wide
429 association study of polymorphisms predisposing to bronchiolitis. *Scientific reports*, 7(1):1–9, 2017.
- 430 [18] David Golan, Eric S Lander, and Saharon Rosset. Measuring missing heritability: inferring the
431 contribution of common variants. *Proceedings of the National Academy of Sciences*, 111 (49):E5272–
432 E5281, 2014.
- 433 [19] AliReza Eshaghi, Venkata R Duvvuri, Rachel Lai, Jeya T Nadarajah, Aimin Li, Samir N Patel, Donald
434 E Low, and Jonathan B Gubbay. Genetic variability of human respiratory syncytial virus a strains
435 circulating in ontario: a novel genotype with a 72 nucleotide g gene duplication. *PloS one*, 7(3):e32807,
436 2012.
- 437 [20] Caroline Breese Hall, Joyce M Geiman, Robert Biggar, David I Kotok, Patricia M Hogan, and R
438 Gordon Douglas Jr. Respiratory syncytial virus infections within families. *New England journal of
439 medicine*, 294(8):414–419, 1976.
- 440 [21] S Levine, R Klaiber-Franco, and PR Paradiso. Demonstration that glycoprotein g is the attachment
441 protein of respiratory syncytial virus. *Journal of General Virology*, 68(9): 2521–2524, 1987.

- 442 [22] Steven A Feldman, R Michael Hendry, and Judy A Beeler. Identification of a linear heparin binding
443 domain for human respiratory syncytial virus attachment glycoprotein g. *Journal of virology*, 73(8):6610–
444 6617, 1999.
- 445 [23] Steven A Feldman, Susette Audet, and Judy A Beeler. The fusion glycoprotein of human respiratory
446 syncytial virus facilitates virus attachment and infectivity via an interaction with cellular heparan sulfate.
447 *Journal of Virology*, 74(14):6442–6447, 2000.
- 448 [24] HW McL Rixon, G Brown, JT Murray, and RJ Sugrue. The respiratory syncytial virus small
449 hydrophobic protein is phosphorylated via a mitogen-activated protein kinase p38- dependent tyrosine
450 kinase activity during virus infection. *Journal of General Virology*, 86 (2):375–384, 2005.
- 451 [25] Reena Ghildyal, Dongsheng Li, Irene Peroulis, Benjamin Shields, Phillip G Bardin, Michael N Teng,
452 Peter L Collins, Jayesh Meanger, and John Mills. Interaction between the respiratory syncytial virus g
453 glycoprotein cytoplasmic domain and the matrix protein. *Journal of General Virology*, 86(7):1879–1884,
454 2005.
- 455 [26] Peter L Collins and Geneviève Mottet. Oligomerization and post-translational processing of
456 glycoprotein g of human respiratory syncytial virus: altered o-glycosylation in the presence of brefeldin a.
457 *Journal of General Virology*, 73(4):849–863, 1992.
- 458 [27] Monica E Brint, Joshua M Hughes, Aditya Shah, Chelsea R Miller, Lisa G Harrison, Elizabeth A
459 Meals, Jacqueline Blanch, Charlotte R Thompson, Stephania A Cormier, and John P DeVincenzo.
460 Prolonged viral replication and longitudinal viral dynamic differences among respiratory syncytial virus
461 infected infants. *Pediatric research*, 82(5): 872–880, 2017.
- 462 [28] Sara M Johnson, Beth A McNally, Ioannis Ioannidis, Emilio Flano, Michael N Teng, Antonius G
463 Oomens, Edward E Walsh, and Mark E Peebles. Respiratory syncytial virus uses cx3cr1 as a receptor on
464 primary human airway epithelial cultures. *PLoS pathogens*, 11(12):e1005318, 2015.
- 465 [29] Ralph A Tripp, Les P Jones, Lia M Haynes, HaoQiang Zheng, Philip M Murphy, and Larry J
466 Anderson. Cx3c chemokine mimicry by respiratory syncytial virus g glycoprotein. *Nature immunology*,
467 2(8):732–738, 2001.
- 468 [30] Kwang-II Jeong, Peter A Piepenhagen, Michael Kishko, Joshua M DiNapoli, Rachel P Groppo,
469 Linong Zhang, Jeffrey Almond, Harry Kleanthous, Simon Delagrange, and Mark Parrington. Cx3cr1 is
470 expressed in differentiated human ciliated airway cells and co-localizes with respiratory syncytial virus on
471 cilia in a g protein-dependent manner. *PloS one*, 10(6):e0130517, 2015.
- 472 [31] Alexander Bukreyev, Lijuan Yang, Jens Fricke, Lily Cheng, Jerrold M Ward, Brian R Murphy, and
473 Peter L Collins. The secreted form of respiratory syncytial virus g glyco- protein helps the virus evade
474 antibody-mediated restriction of replication by acting as an antigen decoy and through effects on fc
475 receptor-bearing leukocytes. *Journal of virology*, 82(24):12191–12204, 2008.
- 476 [32] Megan E Schmidt and Steven M Varga. Modulation of the host immune response by respiratory
477 syncytial virus proteins. *Journal of Microbiology*, 55(3):161–171, 2017.
- 478 [33] Tatiana Chirkova, Christian Rosas-Salazar, Tebeb Gebretsadik, Samadhan J Jadhao, James D
479 Chappell, R Stokes Peebles Jr, William D Dupont, Dawn C Newcomb, Sergejs Berdnikovs, Peter J
480 Gergen, et al. Effect of infant rsv infection on memory t cell responses at age 2-3 years. *Frontiers in
481 immunology*, 13:826666, 2022.
- 482 [34] Andrew R Connelly, Brian M Jeong, Mackenzie E Coden, Jacob Y Cao, Tatiana Chirkova, Christian
483 Rosas-Salazar, Jacqueline-Yvonne Cephus, Larry J Anderson, Dawn C Newcomb, Tina V Hartert, et al.
484 Metabolic reprogramming of nasal airway epithelial cells following infant respiratory syncytial virus
485 infection. *Viruses*, 13(10):2055, 2021.
- 486
- 487

488 **Tables and Figures**

489 **Table 1:** Characteristics of infants with prolonged RSV infection compared with other RSV infection and
490 the entire cohort. Prolonged infection is defined as repeatedly RSV PCR-positive with ≥ 15 days between
491 testing and meeting criteria for acute respiratory infection. *Presence of sibling or another child ≤ 6 years
492 of age at home.

493

494 **Figure 1:** Genetic analyses of RSV infection in infancy. (A) The Manhattan plot shows no genome-wide
495 significant associations (p value threshold of $5e-8$). (B) The Q-Q plot demonstrates that the observed p
496 values are congruent with those expected under the null hypothesis that RSV infection in infancy is
497 independent of host genotype. (C) The association between the 54 selected childhood asthma- or RSV
498 LRTI-associated SNPs and RSV infection in infancy in our data. The identity line is shown in red, and the
499 dashed grey lines are ± 1 standard deviation around the expected $-\log_{10}(p \text{ value})$. RSV: respiratory
500 syncytial virus; SNP: single nucleotide polymorphism.

501

502 **Figure 2:** Viral population structure. (A) Linear map of the RSV genome. (B) Phylogenetic tree based on
503 multiple sequence alignment MSA of G protein amino acid sequences. Color: amino acids. (C) Principal
504 component (PC) analysis. PCs1-3 with labels indicating prolonged infections from different phylogenetic
505 clades. (D) Panel [i] summarises every pairwise genetic distance between every viral sequence. Genetic
506 invariance in prolonged infections separated by at least 15 days compared to other genetic variation
507 within the most closely related sequences (panel [ii]) and within all possible closely related pairs (panel
508 [iii]). VE (variance explained). Jitter applied for visualisation.

509

510 **Figure 3:** Viral genetic association with prolonged infection. (A) Amino acid association with prolonged
511 infection after multiple testing correction (significant threshold shown by dotted line). (B) Variance
512 explained (VE) within cohort. The effect of each variant on cohort structure is shown for PCs1-2. The
513 small % VE for a significantly associated lead variant supports a true positive. (C) Variants in strong
514 correlation were clumped for association testing using proxies for $r^2 \geq 0.8$. One significant association was
515 identified (shown in A); the r^2 values for all other variants show a single highly correlated variant with the
516 lead proxy (red), identifying p.E123K/D and p.P218T/S/L. (D) Evidence for biological interpretation for
517 every amino acid position is summarised. Dotted red lines indicate the positions at p.123, p.218.

518

519 **8 Supplemental methods**

520 **8.1 Study population**

521 The protocol and informed consent documents were approved by the Institutional Review Board at
522 Vanderbilt University Medical Center (#111299). One parent of each participant in the cohort study
523 provided written informed consent for participation in this study. The informed consent document
524 explained study procedures and use of data and biospecimens for future studies, including genetic
525 studies.

526

527 The study population is a longitudinal birth cohort, the INfant Susceptibility to Pulmonary Infections and
528 Asthma Following RSV Exposure (INSPIRE), specifically designed to capture the first RSV infection in
529 term healthy infants. Additional details of this birth cohort have been previously published [41]. Briefly, the
530 cohort included 1949 term (≥ 37 weeks gestation), non-low birth weight (≥ 2250 g, 5 lbs), otherwise
531 healthy infants from a population-representative sample of pediatric practices located in rural, suburban,
532 and urban regions of the south-eastern US during 2012-2014. Infants were born June through December;
533 per study design, they were 6 months of age or less entering their first RSV season.

534

535 **8.2 Biweekly surveillance of RSV infection**

536 Infant (i.e., first year of life) RSV infection was ascertained through passive and active biweekly
537 surveillance during each infant's first RSV season and RSV serology (Table 1). If an infant met pre-
538 specified criteria for an acute respiratory infection, we conducted an in-person respiratory illness visit at
539 which time we administered a parental questionnaire, performed a physical exam, collected a nasal wash,
540 and completed a structured medical chart review for infants seen during an unscheduled visit. RSV RNA
541 in nasal samples was detected by reverse-transcription quantitative PCR [42]. We *a priori* defined the
542 clinical entity of "prolonged" infection during infancy as repeatedly meeting pre-specified criteria for an
543 acute respiratory infection accompanied by repeatedly positive RSV PCR separated by 15 or more days
544 (Figure S1) [13].

545

546 **8.3 Descriptive analyses**

547 Descriptive analyses of the cohort were conducted using R 4.0.5. Pearson or Wilcoxon tests were used
548 for comparing infants with and without prolonged RSV infection. The main descriptive features are
549 provided in Table 1.

550

551 **8.4 Host DNA collection and genotyping**

552 One-year blood samples were selected based on availability of DNA among a subset of children with
553 RSV infection and a random group of those without infection, and were genotyped with the Multi-Ethnic

554 Global Array microarray (Illumina, CA, United States) at the University of Washington DNA Sequencing
555 and Gene Analysis Center (Seattle, WA, United States).

556

557 8.5 Host genetic analyses of RSV infection in infancy

558 To determine whether host genetic factors are associated with infant RSV infection risk, we examined
559 single nucleotide polymorphisms (SNPs) previously shown to alter infant RSV infection severity or
560 childhood asthma risk [15-17]. We also conducted a host GWAS to identify common variants associated
561 with infant RSV infection, and examined narrow sense heritability to test for small cumulative effects. The
562 GWAS was performed on 621 children with available DNA for the association between host genotype and
563 RSV infection during infancy. Due to sample size constraints, we restricted our sub-analysis to the 54
564 host SNPs previously associated with RSV lower respiratory tract infection or childhood asthma [15-17].
565 We additionally evaluated the accumulation of small genetic effects that would go undetected in a GWAS
566 by estimating the narrow sense heritability of RSV infection.

567

568 For GWAS analyses, the initial round of data quality control was performed on individual populations
569 (self-reported as White, Black, and Hispanic) using PLINK version 1.9 [40]. Subjects with a missing
570 genotype call rate above 5% were removed. The SNP minor allele frequency (MAF) threshold was set at
571 > 0.01, 0.03, and 0.08 for White, Black, and Hispanic, respectively [36]. The groups were merged for a
572 total of 1,086,830 variants and a genotyping rate of 0.78. Subject independence was assessed using
573 KING (<https://people.virginia.edu/~wc9c/KING/>) to prevent spurious associations. However, no probable
574 relatives or duplicates were detected based on pairwise identify-by-state. We compared the genetic
575 ancestry in cases to self-reported ethnicity to check for mislabelling. Reported and estimated sex was
576 also examined for discrepancy. A second round of quality control on the combined dataset was
577 conducted, which removed 74 samples due to genotype missingness and 399,991 variants with a
578 genotyping rate < 0.1. Variants were checked for departure from Hardy-Weinberg equilibrium (HWE) ($P <$
579 1e-6) to uncover features of selection, population admixture, cryptic relatedness, or genotyping error. This
580 was only performed on controls to prevent removal of genuine genetic associations that can be
581 associated with this measurement; 6,024 variants were removed. No variants had a MAF < 0.01 after
582 merging. SNP positions and identifiers were compared and updated according to dbNSFP4.0a (hg19)
583 with 289 variants removed due to a missing coordinate and SNPs identifier [35]. This resulted in an
584 analysis-ready dataset of 680,526 variants from 621 children (509 with and 112 without RSV infection in
585 infancy), yielding a total genotyping rate of 0.98. No genomic inflation was evident with an estimated
586 lambda (based on median chi-squared test) equal to 1. We then used genome-wide complex trait
587 analysis (GCTA) software (<https://cnsgenomics.com/software/gcta/>) to calculate the genetic relationship
588 matrix and performed principal component analysis (PCA) to account for population structure [36].
589 Genome-wide association analysis was performed using PLINK version 1.9 for logistic regression with
590 multiple covariates consisting of the child's birth month, enrolment year (as a marker of RSV season),

591 daycare attendance, presence of another child \leq 6 years of age at home, sex, and 6 ancestry principal
592 components (PCs) [40].

593

594 As the multiple testing burden likely precluded identification of small genetic effects in our GWAS, we
595 conducted an additional heritability analysis using the method described by Golan et al. [18] to estimate
596 narrow-sense heritability of RSV infection during infancy on the latent liability scale (h^2_l), which, if > 0 ,
597 would indicate an accumulation of small genetic effects. We estimated h^2_l to be exactly 0, suggesting that,
598 if present, infant RSV infection-related genetic signals are both small and sparse.

599

600 8.6 Host genetic analyses for known associations

601 We further investigated the possibility that the analysis was underpowered to identify associations with
602 reported childhood asthma- and RSV LRTI-associated SNPs [15-17]. This was done by pooling
603 information across SNPs to estimate the average genetic effect size. In brief, we computed a z-score for
604 each SNP, where the average (across SNPs) squared z-score $G(\bar{z})$ is proportional to the average
605 squared genetic effect on RSV infection in infancy. As $G(\bar{z})$ is an average of $p = 54$ approximately
606 independent statistics, it is approximately $N(n\mu^2 + 1, 2/p)$, where $n = 621$ is the sample size and μ^2 is a
607 function of the average squared genetic effect on RSV infection in infancy. Using the genetic effect
608 estimates from Pividori et al. [15], Janssen et al. [16], and Pasanen et al. [17], we calculated that we
609 would have 80% power to reject the global null hypothesis of no genetic effect at any of these SNPs (i.e.,
610 $\mu^2 = 0$) if, on average across the 54 SNPs, the genetic effect on RSV infection in infancy was at least 61%
611 as large as those estimated in the aforementioned 3 studies. We found $G(\bar{z})=1.00$ in our data, which
612 corresponds to a p value of 0.50. This result indicates that the genetic effect on RSV infection in infancy is
613 zero or small at SNPs likely to be associated with RSV infection *a priori*.

614

615 8.7 Host acute local immune response

616 Nasal wash samples collected at the time of acute infant RSV infection were profiled to measure the
617 acute host response to infection using Luminex xMap multianalyte bead assays (Milliplex Human
618 Cytokine/Chemokine Panel II MAGNETIC Premixed 23 Plex Kit, EMD Millipore; and Cytokine 30-Plex
619 Human Panel, Life Technologies Corporation). These data were used to test the host nasal interferon
620 (IFN) response and the viral variant associated with prolonged infection.

621

622 8.8 RSV whole-genome sequencing

623 RSV genome sequencing was performed on all specimens from subjects meeting illness criteria and with
624 positive RSV PCR. Viral amino acid variants (genotype) of the F and G glycoprotein were tested for
625 association with prolonged infection adjusting for host features associated with increased infection risk.
626 The relatively small sample size of our cohort required analysis that targeted only genes which were a

627 priori likely to functionally contribute to the clinical phenotype. Therefore, our analysis focused on the
628 surface F (fusion) and G (attachment) proteins of RSV as they have been implicated in pathogenesis [43;
629 44], and both are targets for neutralizing antibodies during infection [45; 46]. Lastly, to determine if the
630 variants of interest were enriched by selective pressure over time, we used public data from the past
631 several decades to assess variant frequency over time.

632

633 RSV whole-genome sequencing of this study population has been previously described [47]. Briefly, RNA
634 was extracted at J. Craig Venter Institute (JCVI) (<https://www.jcvi.org>) in Rockville, MD from nasal wash
635 samples which were RSV PCR positive and collected during a respiratory illness visit triggered through
636 biweekly surveillance of symptoms. Four forward reverse-transcription (RT) primers were designed and
637 four sets of PCR primers were manually picked across a consensus of complete RSV genome sequences
638 using JCVI's automated primer design tool [48]. cDNA was generated from 4 µL undiluted RNA using the
639 pooled forward primers and SuperScript III Reverse Transcriptase (Thermo Fisher Scientific, Waltham,
640 MA, USA). 100 ng of pooled DNA amplicons were sheared to create 400-bp libraries, which were pooled
641 in equal volumes and cleaned with Ampure XP reagent (Beckman Coulter, Inc., Brea, CA, USA).
642 Sequencing was performed on the Ion Torrent PGM using 316v2 or 318v2 chips (Thermo Fisher
643 Scientific).

644

645 For samples requiring extra coverage, in addition to Ion Torrent sequencing, Illumina libraries were
646 prepared using the Nextera DNA Sample Preparation Kit (Illumina, Inc., San Diego, CA, USA). Sequence
647 reads were sorted by barcode, trimmed, and assembled de novo using CLC Bio's *clc novo assemble*
648 program, and the resulting contigs were searched against custom, full-length RSV nucleotide databases
649 to find the closest reference sequence. All sequence reads were then mapped to the selected reference
650 RSV sequence using CLC Bio's *clc ref assemble long* program [49]. Curated assemblies were validated
651 and annotated with the viral annotation software called Viral Genome ORF Reader, VIGOR 3.0
652 (<https://sourceforge.net/projects/jcvi-vigor/files/>), before submission to GenBank as part of the Bioproject
653 accession PRJNA225816 (<https://www.ncbi.nlm.nih.gov/bioproject/225816>) [50] and PRJNA267583
654 (<https://www.ncbi.nlm.nih.gov/bioproject/267583>).

655

656 8.9 Viral sequence alignment

657 The NCBI-tools, Tbl2asn (<https://www.ncbi.nlm.nih.gov/genbank/tbl2asn2/>), was used in the creation of
658 sequence records for submission to GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). A total of 350
659 viral sequences in .sqn file format were used for downstream analysis.

660

661 We computed a phylogenetic tree for each gene, as follows. NCBI-tools asn2fsa (<https://www.huge-man->
662 <https://linux.net/man1 asn2fsa.html>) was used to convert sequences to fasta format. Each sample consisted of
663 11 sequence segments (NS1, NS2, N, P, M, M2-1, M2-2, SH, G, F, and L) as shown in Figure 1. These

664 were separated and repooled to create 11 single fasta files for each gene containing all 350 samples.
665 Sequences were checked for at least 90% coverage of the corresponding gene to minimize loss of
666 aligned positions when computing the phylogenetic tree. Each of the eleven resulting sets was aligned
667 with MAFFT v7 (<https://mafft.cbrc.jp/alignment/software/>) [39], using default parameters. The sequence of
668 the orthologous gene from Bovine orthopneumovirus ([GenBank:NC 001989](#)) was added to each set as an
669 outgroup.

670

671 IQ-Tree (<https://www.iqtree.org>) [37] was used with per-gene multiple sequence alignment (MSA) files
672 based on amino acid sequence for estimating maximum-likelihood phylogenies using protein substitution
673 model. Examining the sequences with an alignment viewer showed that a small number of sequences
674 had frame-shift variants but these did not affect the regions included in our testing criteria.

675

676 Viral sequence data and clinical information were merged and cleaned with R. Clinical IDs matching more
677 than one viral sequence ID were used to re-identify samples from the same individual as prolonged
678 infections. Genetic variation was quantified in these samples, and for subsequent analysis, only the first
679 viral sequence was included for association testing. Antigenic grouping of strain A and B had been
680 completed previously and labels were included to annotate each sample accordingly.

681

682 The cohort-specific variant frequency per position was calculated; residues were counted and ranked by
683 frequency with the most frequent residue defined as reference (REF) and alternative (ALT) for variants.
684 Positions with at least one ALT were checked for potential misalignment or other sources of error. Variant
685 positions were selected for association analysis, while non-variant positions were ignored.

686

687 A number of host features have been previously shown to influence infection susceptibility and were
688 therefore included as covariates in our analysis [51]. Six samples were excluded due to insufficient
689 covariate data, resulting in 344 test samples. Of these, 38 were from the same patients (prolonged
690 infection) of which half (19) were included for association testing. Thus, the test set was comprised of
691 single samples collected from 325 individuals.

692

693 8.10 Viral population structure

694 The genetic distances to nearest neighbors were computed based on phylogenetic trees generated with
695 MAFFT. PCA and singular value decomposition (SVD) were used in dimensionality reduction for
696 exploratory data analysis of viral phylogeny. The R package factoextra was used for PCA and to visualise
697 eigenvalues and variance. R package caret was used to analyse genetic correlations.

698

699 8.11 Viral variant association testing

700 Viral amino acids (genotype collapsed into REF/ALT) were tested for association with infection types (i.e.,
701 resolved and prolonged) including key covariates that alter infection risk. To reduce the multiple testing
702 burden, proxy amino acid variants were identified by performing clumping with ranking based on MAF and
703 with a cut-off threshold of $r^2 \geq 0.8$ (Supplemental Figure S3). Since many variants within RSV coding
704 genes have non-random association due to selection, analogous to linkage disequilibrium (LD) in human
705 GWAS, we reduced the multiple testing burden by retaining proxy variants and removing those with $r^2 \geq$
706 0.8. Analysis was performed using logistic regression with the R stats (3.6.2) glm function. The model
707 consisted of the binary response (prolonged infection Yes/No) and predictors viral genotype (REF/ALT
708 amino acid, including multi-allelic non-REF collapsed into ALT), viral PCs 1-5, host sex, and host features
709 that have been previously demonstrated as significantly associated with infection: daycare attendance,
710 living with siblings and self-reported race/ethnicity [51].

711

712 Environmental host covariates did not contribute significant effect in our model for candidate causal
713 association. For this reason, in our main analysis, viral population structure was accounted for by the first
714 five PCs. The Bonferroni correction for multiple testing was applied based on the number of variants
715 tested. For the significant association found by proxy amino acid variants, the association test was
716 repeated for all clumped variants to produce a LocusZoom-style Manhattan plot containing r^2 by color and
717 p value statistics. R package stats was used for a range of analyses including glm for logistic regressions.
718 R package MASS was used to analyse logistic regression model data. To test if the significantly
719 associated variants were due to population structure, we re-estimated models using the subset of
720 individuals infected with RSV strain B to confirm validity of combined analysis.

721

722 8.12 Public viral sequence data

723 We gathered publicly available sequence data to further assess variants of interest. We used the public
724 viral data repository of NCBI https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Human%20orthopneumovirus,%20taxid:11250 to
725 retrieve information using search criteria that follow. Virus: Human orthopneumovirus (HRSV),
726 taxid:11250. Proteins: attachment glycoprotein. Host: Homo (humans), taxid:9605. Collection dates: Jan
727 1, 1956 onward. Nucleotide and protein sequence data was collected, which contained data from 27
728 countries and 1084 glycoprotein protein sequences after curation. Sequence and meta data were
729 merged. Multiple sequence alignment was performed to find consensus relative positions for all
730 sequences. Regions of interest were then extracted and re-annotated with their correct amino acid
731 positions matching the reference sequence. Summary statistics were generated, including number of
732 samples, collection date, geo-location, variant frequency, and strain for the specified amino acid
733 (Supplemental Figure S4).

735

736 8.13 Biological interpretation

737 As infant RSV infection stimulates an acute antiviral response and also results in decreased barrier
738 function of the airway epithelium [34], we tested for association between host interferon (IFN) response
739 and the amino acid (REF/ALT) identified as the viral variant associated with prolonged infection. A
740 Wilcoxon test was performed to compare IFN- γ and IFN- α between RSV amino acid positions, with
741 adjustment for the same covariates as in the main analysis. Protein structures were analysed with data
742 sourced from RCSB PDB <https://www.rcsb.org>. Protein function and domains were assessed using
743 UniProt (<https://www.uniprot.org>) for P03423 (GLYC HRSVA) (strain A2) and O36633 (GLYC HRSVB)
744 (strain B1) in gff format ([https://www.uniprot.org/uniprot/ P03423](https://www.uniprot.org/uniprot/P03423) and
745 <https://www.uniprot.org/uniprot/O36633>, respectively). Interactions, post-translational modifications,
746 motifs, and epitopes were assessed from the literature. Protein features were assessed using data from
747 NCBI (https://www.ncbi.nlm.nih.gov/ipg/NP_056862.1) and via sequence viewer with O36633.1 human
748 RSV B1 ([https://www.ncbi.nlm.nih.gov/projects/sviewer/?id= O36633.1](https://www.ncbi.nlm.nih.gov/projects/sviewer/?id=O36633.1)) Potential effects of amino acid
749 variation on protein structure and function were considered according to available information on a broad
750 range of biological and biochemical features, including native conformation (secondary, tertiary, and
751 quaternary), domains and topology, disulfide bonds, glycosylation, interactions with other viral proteins
752 and host-cell factors, proteolytic cleavage sites, normal patterns of intra-and/or extra-cellular distribution,
753 and secretion status.

754

- 755 **Supplemental references**
- 756 Continued form main text
- 757 [35] Xiaoming Liu, Chunlei Wu, Chang Li, and Eric Boerwinkle. dbnsfp v3. 0: A one-stop database of
758 functional predictions and annotations for human nonsynonymous and splice-site snvs. Human mutation,
759 37(3):235–241, 2016.
- 760 [36] Jian Yang, S Hong Lee, Michael E Goddard, and Peter M Visscher. Gcta: a tool for genome-wide
761 complex trait analysis. *The American Journal of Human Genetics*, 88(1): 76–82, 2011.
- 762 [37] Lam-Tung Nguyen, Heiko A Schmidt, Arndt Von Haeseler, and Bui Quang Minh. Iq-tree: a fast and
763 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and*
764 *evolution*, 32(1):268–274, 2015.
- 765 [38] Ani Manichaikul, Josyf C. Mychaleckyj, Stephen S. Rich, Kathy Daly, Michèle Sale, and Wei-Min
766 Chen. Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22):2867–
767 2873, 2010. Number: 22 ISBN: 1460-2059 Publisher: Oxford University Press.
- 768 [39] Kazutaka Katoh and Daron M Standley. Mafft multiple sequence alignment software version 7:
769 improvements in performance and usability. *Molecular biology and evolution*, 30 (4):772–780, 2013.
- 770 [40] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel AR Ferreira, David
771 Bender, Julian Maller, Pamela Sklar, Paul IW De Bakker, Mark J Daly, et al. Plink: a tool set for whole-
772 genome association and population-based linkage analyses. *The American journal of human genetics*,
773 81(3):559–575, 2007.
- 774 [41] E. K. Larkin, T. Gebretsadik, M. L. Moore, L. J. Anderson, W. D. Dupont, J. D. Chappell, P. A.
775 Minton, R. S. Peebles, Jr., P. E. Moore, R. S. Valet, D. H. Arnold, C. Rosas-Salazar, S. R. Das, F. P.
776 Polack, and T. V. Hartert. Objectives, design and enrollment results from the Infant Susceptibility to
777 Pulmonary Infections and Asthma Following RSV Exposure Study (INSPIRE). *BMC Pulm Med*, 15:45,
778 April 2015. ISSN 1471-2466. doi: 10.1186/s12890-015-0040-0. Edition: 2015/05/30.
- 779 [42] Emma K Larkin, Tebeb Gebretsadik, Martin L Moore, Larry J Anderson, William D Dupont, James D
780 Chappell, Patricia A Minton, R Stokes Peebles, Paul E Moore, Robert S Valet, et al. Objectives, design
781 and enrollment results from the infant susceptibility to pulmonary infections and asthma following rsv
782 exposure study (inspire). *BMC pulmonary medicine*, 15(1):1–12, 2015.
- 783 [43] Seyhan Boyoglu-Barnum, Sean O Todd, Tatiana Chirkova, Thomas R Barnum, Kelsey A Gaston, Lia
784 M Haynes, Ralph A Tripp, Martin L Moore, and Larry J Anderson. An anti-g protein monoclonal antibody
785 treats rsv disease more effectively than an anti-f monoclonal antibody in balb/c mice. *Virology*, 483:117–
786 125, 2015.
- 787 [44] Alexander Bukreyev, Lijuan Yang, and Peter L Collins. The secreted g protein of hu- man respiratory
788 syncytial virus antagonizes antibody-mediated restriction of replication involving macrophages and
789 complement. *Journal of virology*, 86(19):10880–10884, 2012.
- 790 [45] Larry J Anderson, P Bingham, and JC Hierholzer. Neutralization of respiratory syncytial virus by
791 individual and mixtures of f and g protein monoclonal antibodies. *Journal of virology*, 62(11):4232–4238,
792 1988.
- 793 [46] Joan O Ngwuta, Man Chen, Kayvon Modjarrad, M Gordon Joyce, Masaru Kanekiyo, Azad Kumar,
794 Hadi M Yassine, Syed M Moin, April M Killikelly, Gwo-Yu Chuang, et al. Prefusion f-specific antibodies
795 determine the magnitude of rsv neutralizing activity in human sera. *Science translational medicine*,
796 7(309):309ra162–309ra162, 2015.
- 797 [47] S. A. Schobel, K. M. Stucker, M. L. Moore, L. J. Anderson, E. K. Larkin, J. Shankar, J. Bera, V. Puri,
798 M. H. Shilts, C. Rosas-Salazar, R. A. Halpin, N. Fedorova, S. Shrivastava, T. B. Stockwell, R. S.
799 Peebles, T. V. Hartert, and S. R. Das. Respiratory Syncytial Virus whole-genome sequencing identifies
800 convergent evolution of sequence duplication in the C-terminus of the G gene. *Sci Rep*, 6:26311, May
801 2016. ISSN 2045-2322. doi: 10.1038/srep26311. Edition: 2016/05/24.

- 802 [48] K. Li, S. Shrivastava, A. Brownley, D. Katzel, J. Bera, A. T. Nguyen, V. Thovarai, R. Halpin, and T. B.
803 Stockwell. Automated degenerate PCR primer design for high- throughput sequencing improves
804 efficiency of viral sequencing. *Virol J*, 9:261, November 2012. ISSN 1743-422x. doi: 10.1186/1743-422x-
805 9-261. Edition: 2012/11/08.
- 806 [49] QIAGEN Aarhus. White paper on de novo assembly in CLC Assembly Cell 4.0. *digi- talinsights*, page
807 14, June 2016. URL <https://digitalinsights.qiagen.com/files/whitepapers/whitepaper-denovo-assembly.pdf>. Place: Denmark Publisher: Qiagen.
- 809 [50] S. Wang, J. P. Sundaram, and T. B. Stockwell. VIGOR extended to annotate genomes for additional
810 12 different viruses. *Nucleic Acids Res*, 40(Web Server issue):W186–92, July 2012. ISSN 0305-1048
811 (Print) 0305-1048. doi: 10.1093/nar/gks528. Edition: 2012/06/07.
- 812 [51] Christian Rosas-Salazar, Zheng-Zheng Tang, Meghan H Shilts, Kedir N Turi, Qilin Hong, Derek A
813 Wiggins, Christian E Lynch, Tebeb Gebretsadik, James D Chappell, R Stokes Peebles Jr, et al. Upper
814 respiratory tract bacterial-immune interactions during respiratory syncytial virus infection in infancy.
815 *Journal of Allergy and Clinical Immunology*, 149(3): 966–976, 2022.

816

817 **Supplemental figures**

818

819 **Figure S1:** INFant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure
820 (INSPIRE). The study population is a longitudinal birth cohort specifically designed to capture the first
821 RSV infection in term healthy infants. Prolonged infection was a priori defined as repeatedly meeting
822 criteria for acute respiratory infection with RSV PCR positive nasal samples ≥ 15 days between testing.
823

824 **Figure S2:** Supplemental: Infant RSV prolonged infections. Each line represents an infant in the study,
825 and line start and end correspond to clinical respiratory illness sampling timepoints. CT values are
826 inversely related to viral RNA abundance.

827

828 **Figure S3:** Supplemental: Variant clumping for reduction in association testing. [Left] Correlation between
829 all positions. [Right] Correlation between proxy variants after clumping to remove $r_2 \geq 0.8$. Values indicate
830 relative amino acid positions within MSA. r_2 indicated by color scale.

831

832 **Figure S4:** Supplemental: Publicly available RSV sequence data for > 30 years. (A) Global sample
833 collection per year. (B) Variant associated with prolonged infection tracked in public data. The lead proxy
834 SNP, p.P218T/S/L is illustrated here (relative amino acid position 410 in MSA). The major alleles (proline,
835 leucine) are seen for group A/B, with minor alleles (serine, threonine) generally at low frequency <10%.
836 (C) % variance explained per year for all G protein amino acid variants from 1990-2022.

Table 1

	Prolonged RSV infection N = 19	RSV infection N=342	Total N = 1949
Age in months at first RSV illness (median, IQR)	6 (4, 6)	4 (2,5)	NA
Illness respiratory severity score	2.0 (1.2, 3.0)	3.0 (2.0, 4.0)	NA
RSV season			
2012-13	68%	54%	44%
2013-14	32%	46%	56%
Self-reported race			
Non-Hispanic Black	37%	13%	18%
Non-Hispanic White	63%	69%	65%
Hispanic	0%	10%	9%
Multi race/ethnicity/other	0%	8%	8%
Female sex	53%	44%	48%
Second-hand smoke exposure	21%	23%	47%
Health insurance			
Medicaid	68%	48%	54%
Private	32%	51%	45%
None/unknown	0%	1%	1%
Daycare and/or siblings	84%	78%	66%

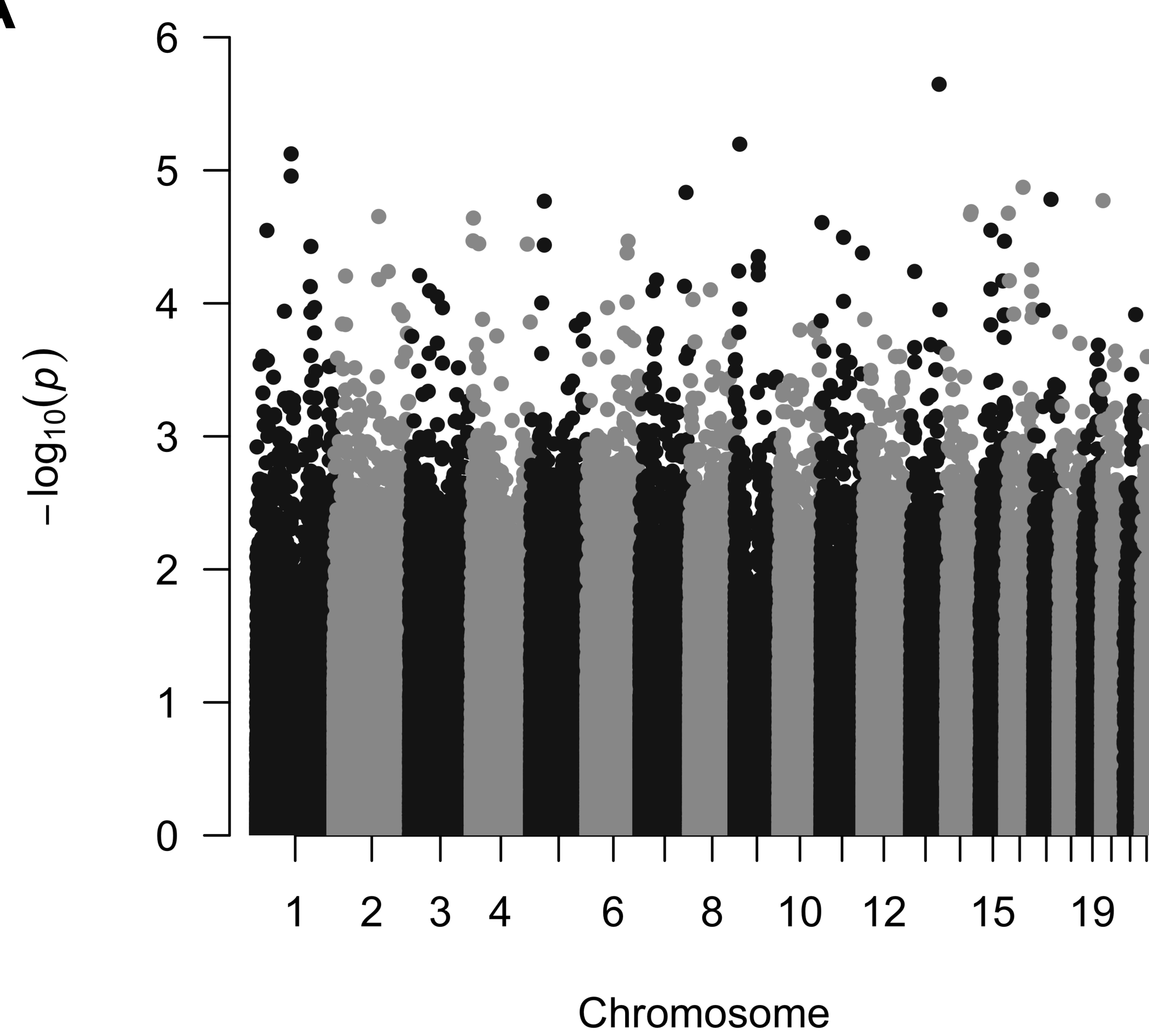
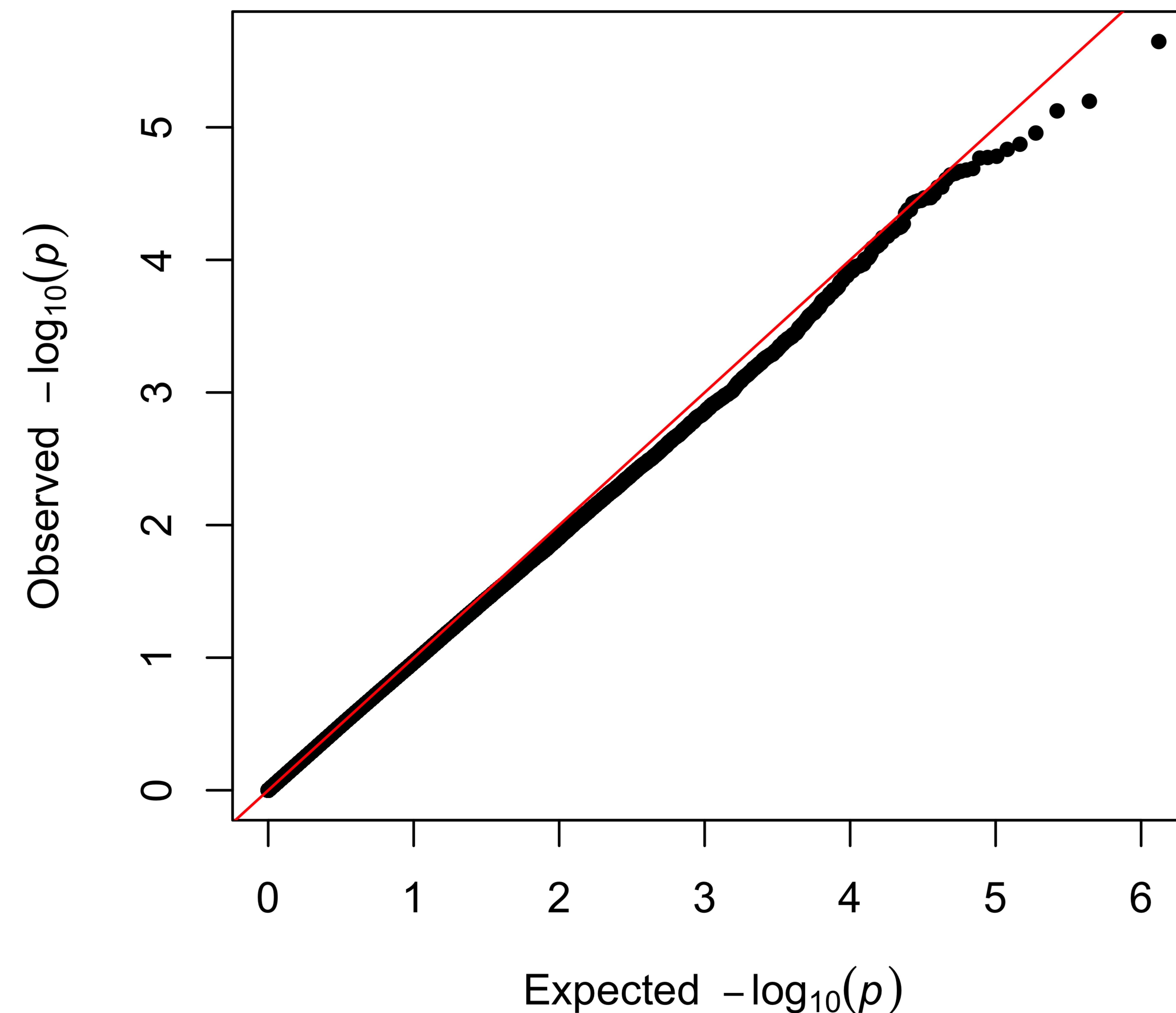
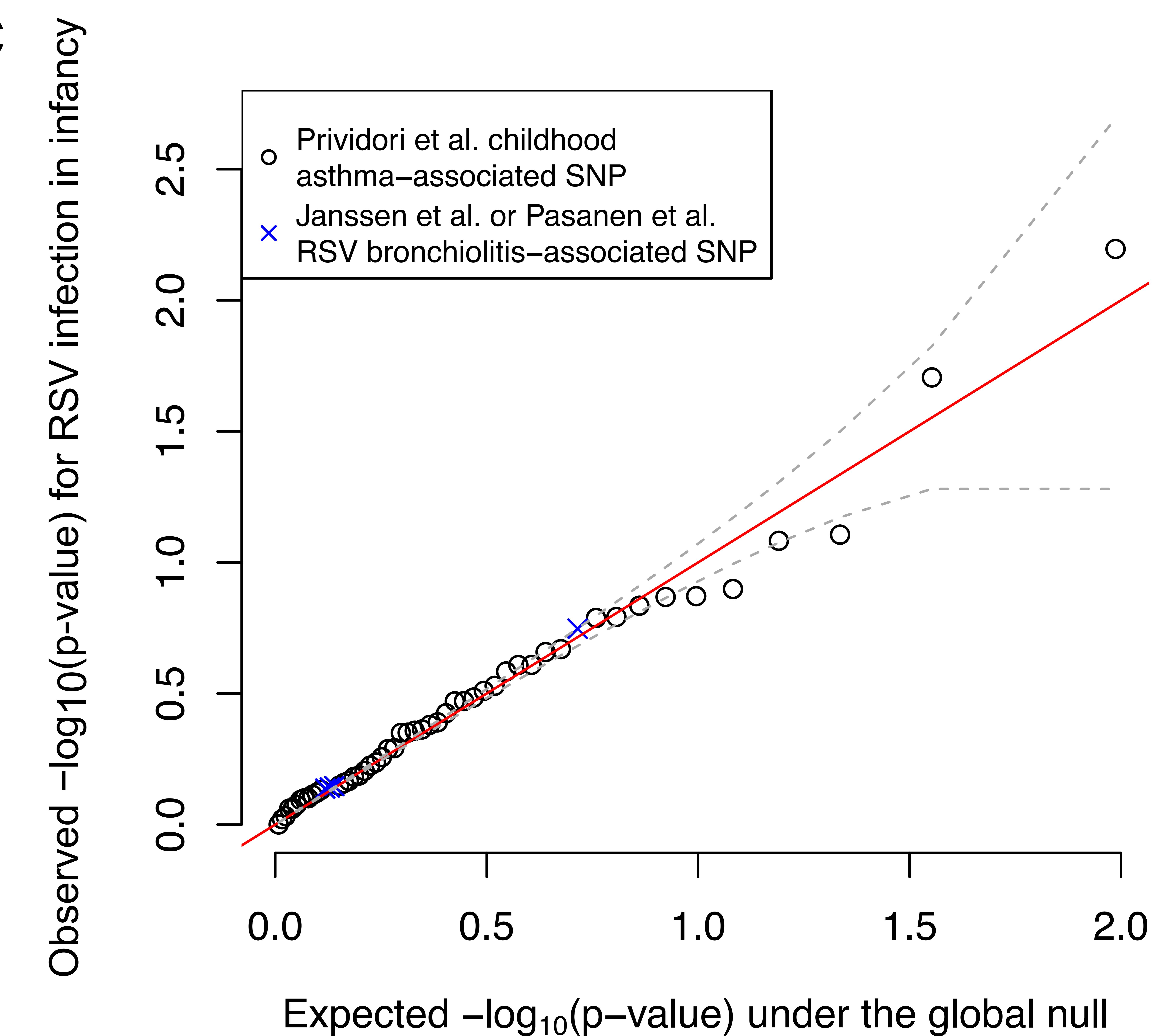
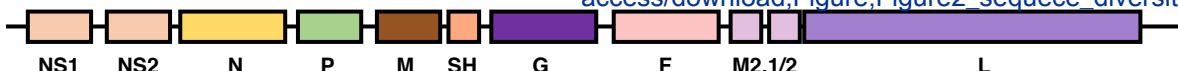
A**B****C**

Figure 2

[Click here to access/download;Figure;Figure2_sequece_diversity.pdf](#)

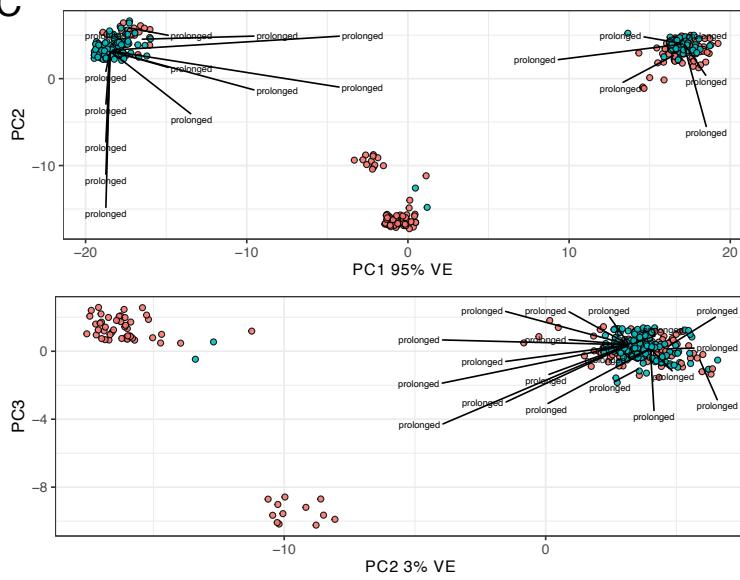
A



B



C



D

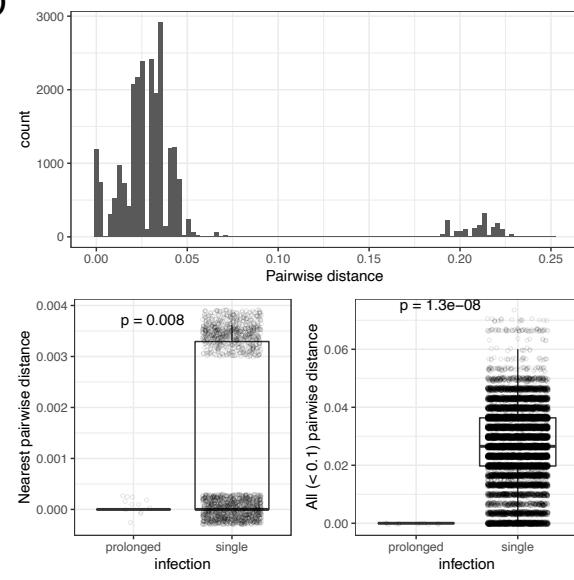
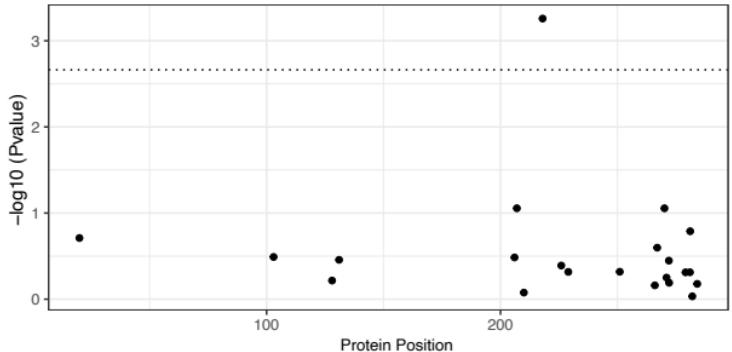
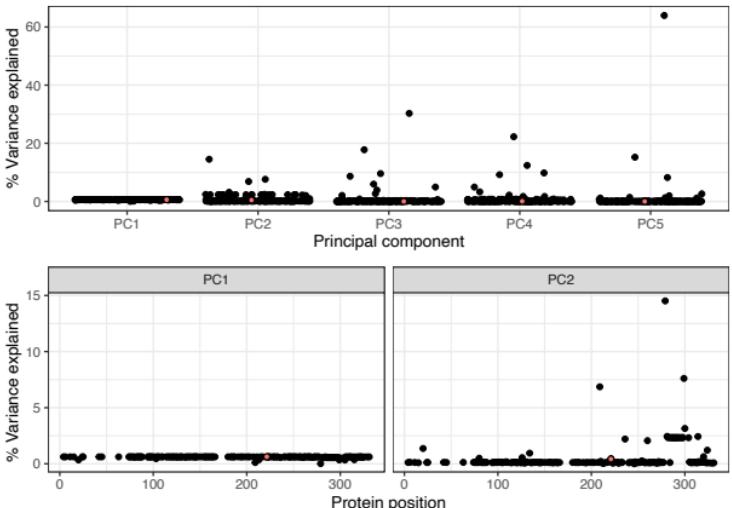


Figure 3

A



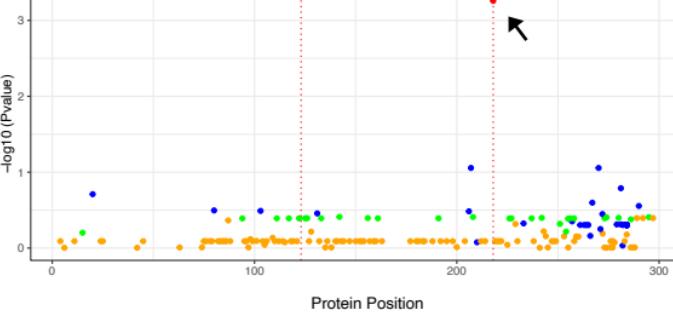
B



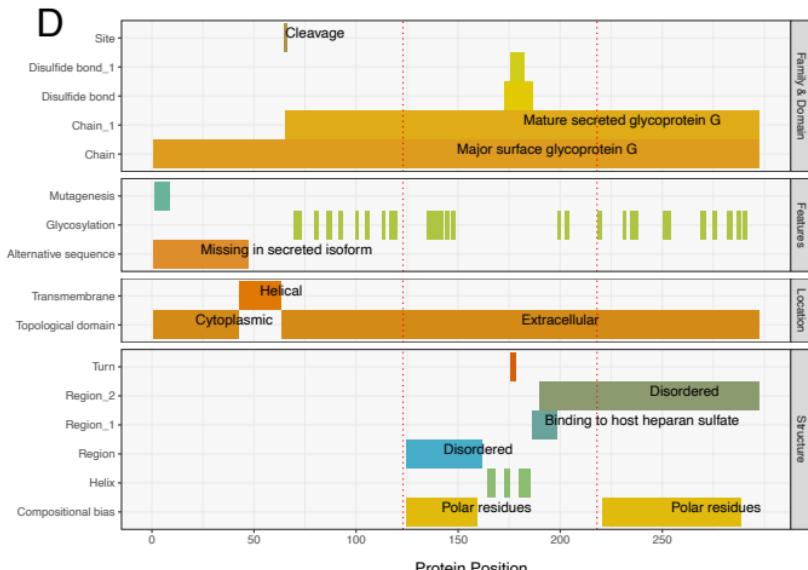
[Click here to access/download Figure 3 viral genetic associations](#)

 r^2 with proxy SNP

- (0,0.4]
- (0.4,0.6]
- (0.6,0.8]
- (0.8,1]



D



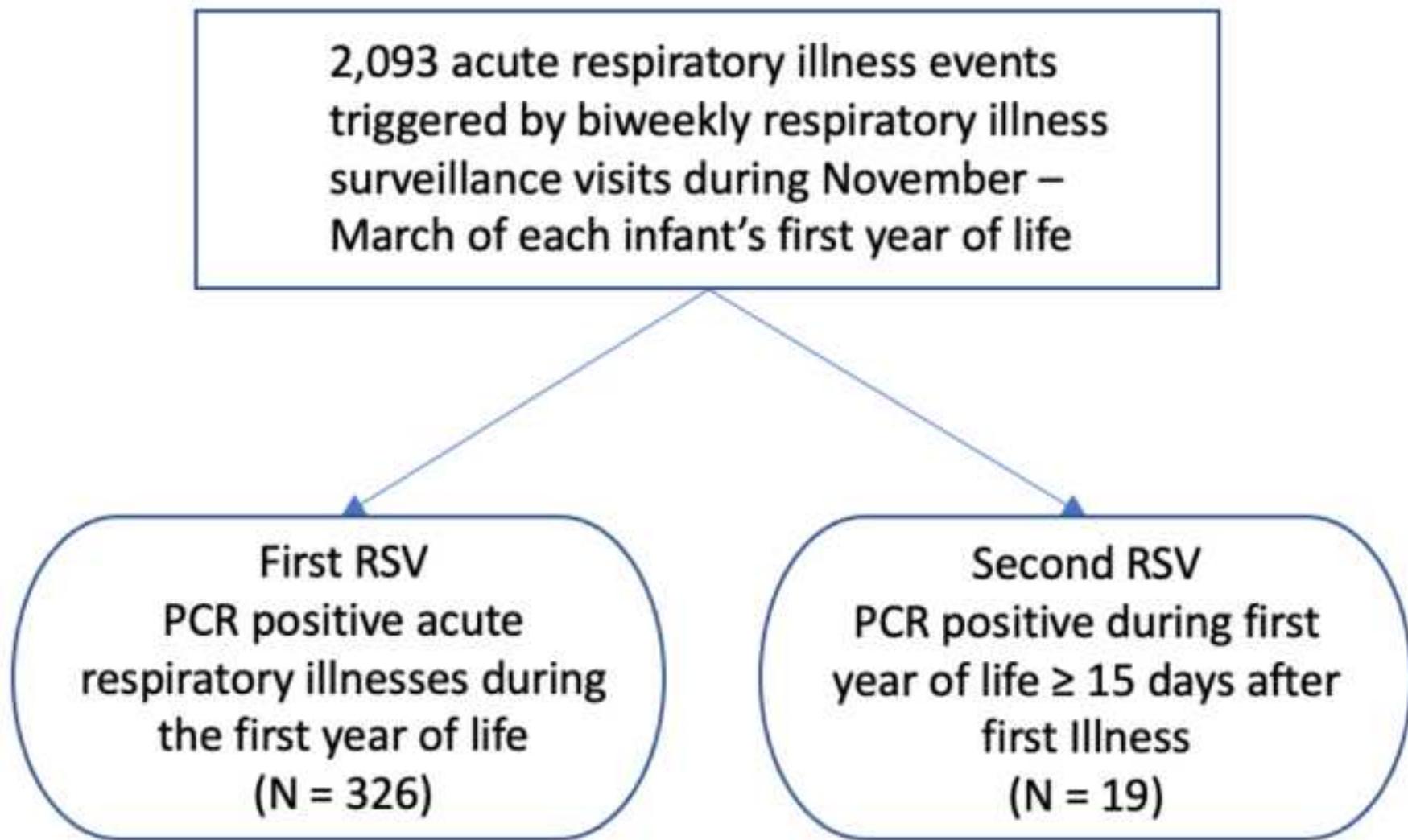


Figure S2

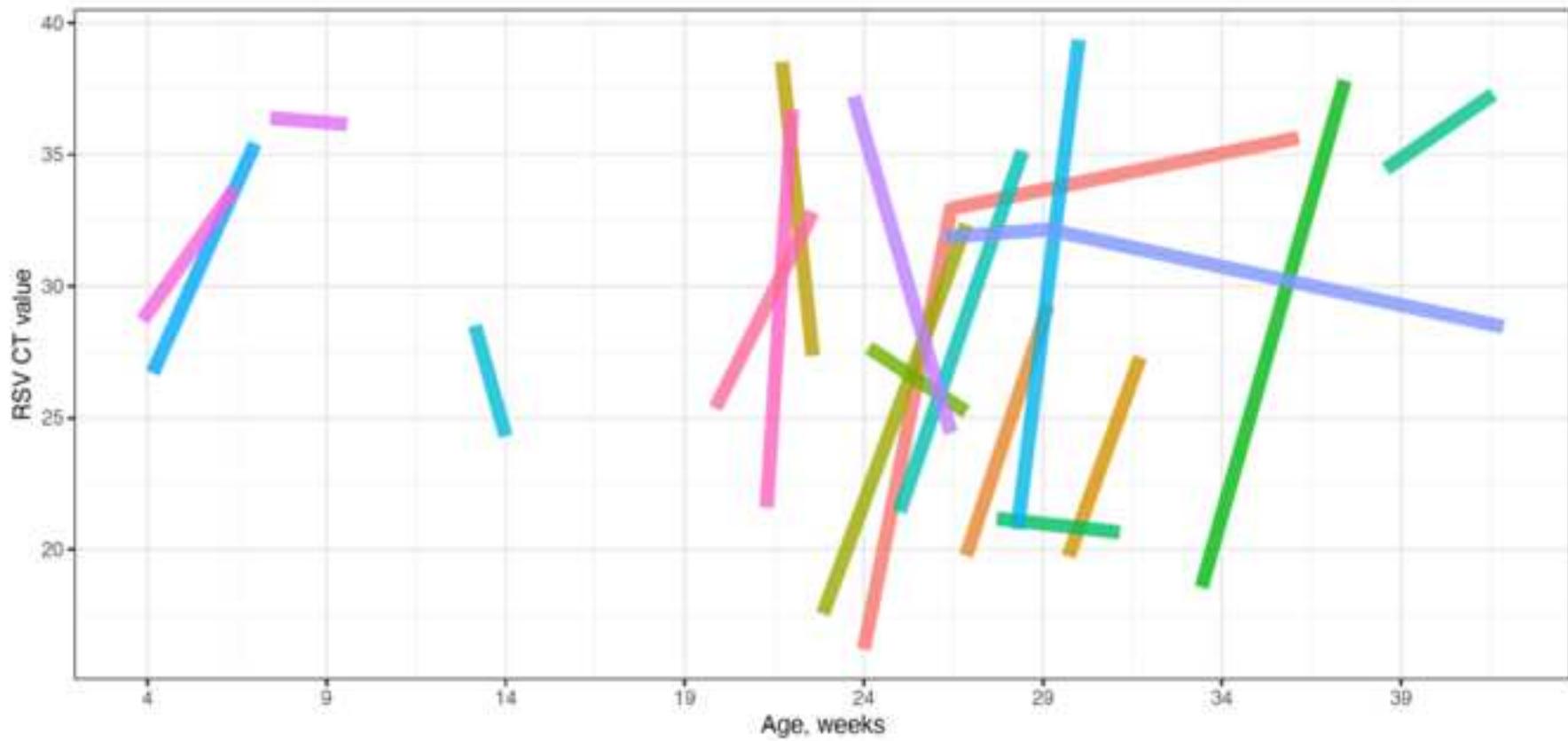
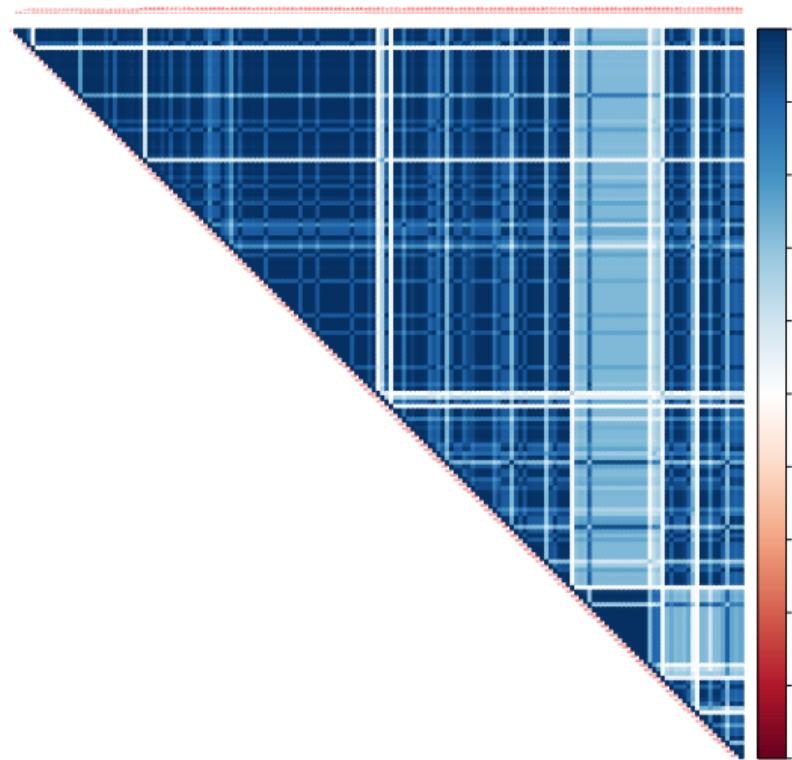
[Click here to access/download;Figure;FigureS2_CT_values.tif](#)

Figure S3

A



B

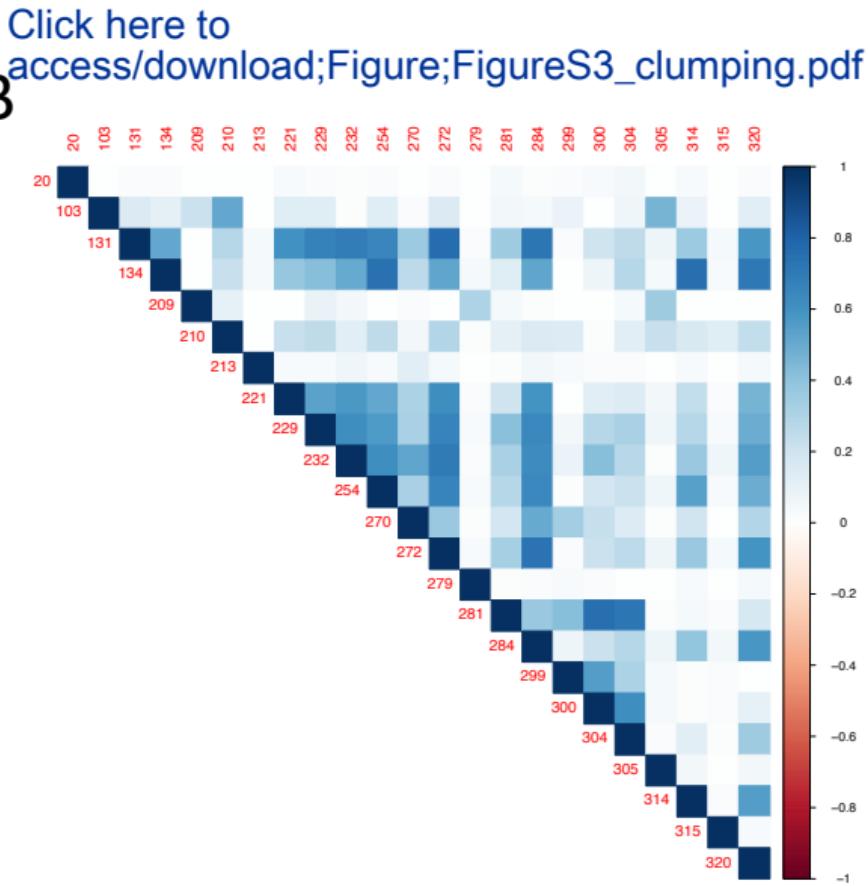
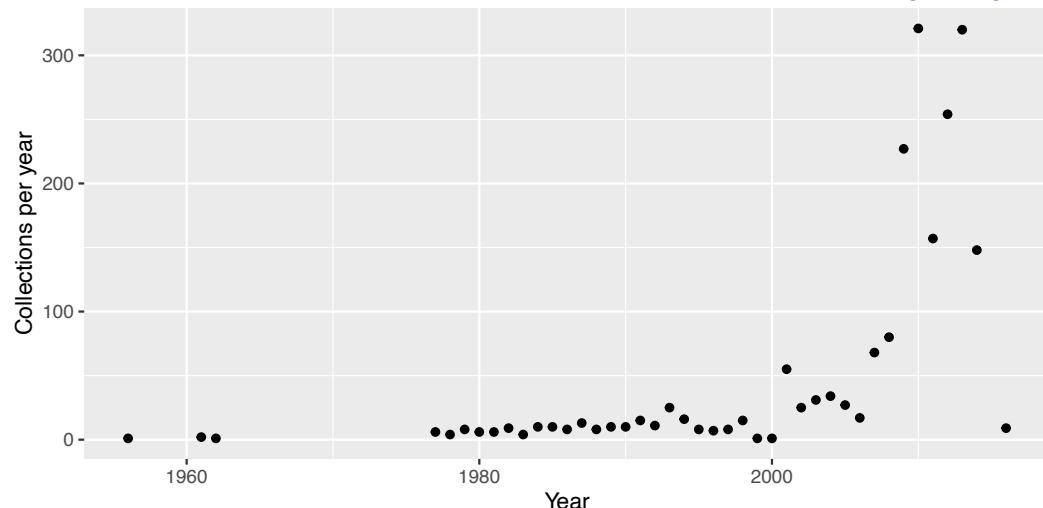
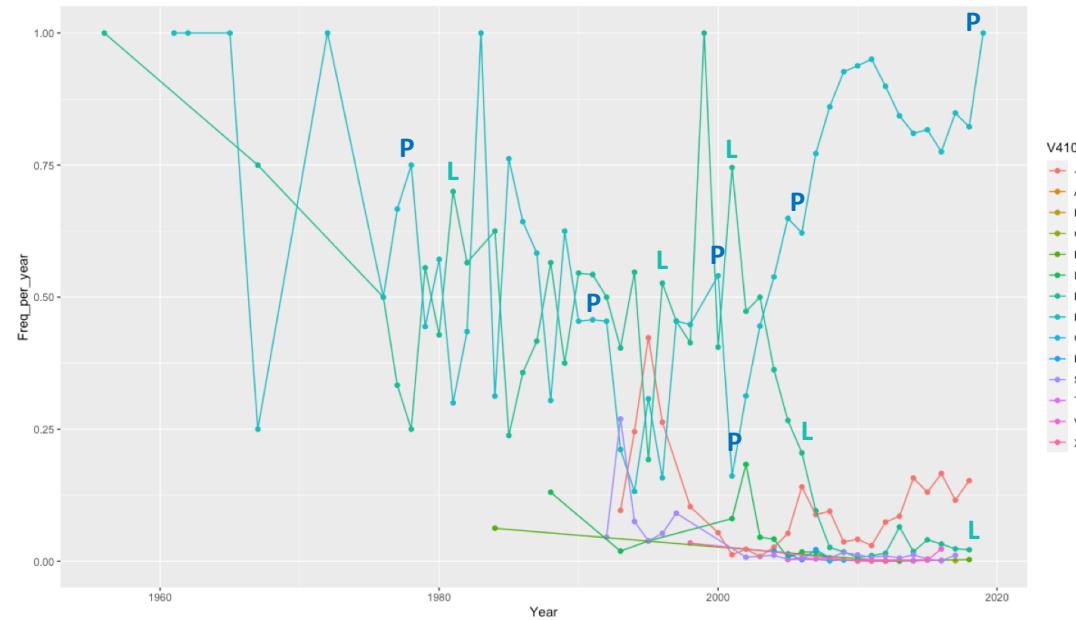


Figure S4
[Click here to access/download;Figure;FigureS4_historical_public_seq](#)

A



B



C

