

1 *Title*

2 **Type 2 and interferon inflammation strongly regulate SARS-CoV-2 related gene**
3 **expression in the airway epithelium**

4

5

6 *Authors/Affiliations*

7 Satria P. Sajuthi^{1#}, Peter DeFord^{1#}, Nathan D. Jackson¹, Michael T. Montgomery¹,
8 Jamie L. Everman¹, Cydney L. Rios¹, Elmar Pruesse¹, James D. Nolin¹, Elizabeth G.
9 Plender¹, Michael E. Wechsler², Angel CY Mak³, Celeste Eng³, Sandra Salazar³, Vivian
10 Medina⁴, Eric M. Wohlford^{3,5}, Scott Huntsman³, Deborah A. Nickerson^{6,7,8}, Soren
11 Germer⁹, Michael C. Zody⁹, Gonçalo Abecasis¹⁰, Hyun Min Kang¹⁰, Kenneth M. Rice¹¹,
12 Rajesh Kumar¹², Sam Oh³, Jose Rodriguez-Santana⁴, Esteban G. Burchard^{3,13}, Max A.
13 Seibold^{1,14,15*}

14

15 ¹Center for Genes, Environment, and Health, National Jewish Health, Denver, CO,
16 80206 USA; ²Department of Medicine, ¹⁴Department of Pediatrics, National Jewish
17 Health, Denver, CO, 80206 USA; ¹⁵Division of Pulmonary Sciences and Critical Care
18 Medicine, University of Colorado-AMC, Aurora, CO, 80045 USA, ³Department of
19 Medicine, ⁵Division of Pediatric Allergy and Immunology, ¹³Department of
20 Bioengineering and Therapeutic Sciences University of California San Francisco, San
21 Francisco, CA; ⁶Department of Genome Sciences, University of Washington, Seattle,
22 WA, USA, ⁷Northwest Genomics Center, Seattle, WA, USA ⁸Brotman Baty Institute,
23 Seattle, WA, USA; ⁹New York Genome Center, NYC, New York; ¹⁰Center for Statistical

24 Genetics, University of Michigan, Ann Arbor, MI, USA; ¹¹Department of Biostatistics,
25 University of Washington, Seattle, WA, USA; ⁴Centro de Neumología Pediátrica, San
26 Juan, Puerto Rico; ¹²Ann and Robert H. Lurie Children's Hospital of Chicago,
27 Department of Pediatrics, Northwestern University, Chicago, IL [#]these authors
28 contributed equally to this work, *correspondence: seiboldm@njhealth.org

29

30

31 *Abstract*

32

33 Coronavirus disease 2019 (COVID-19) outcomes vary from asymptomatic infection to
34 death. This disparity may reflect different airway levels of the SARS-CoV-2 receptor,
35 ACE2, and the spike protein activator, TMPRSS2. Here we explore the role of genetics
36 and co-expression networks in regulating these genes in the airway, through the
37 analysis of nasal airway transcriptome data from 695 children. We identify expression
38 quantitative trait loci (eQTL) for both *ACE2* and *TMPRSS2*, that vary in frequency
39 across world populations. Importantly, we find *TMPRSS2* is part of a mucus secretory
40 network, highly upregulated by T2 inflammation through the action of interleukin-13, and
41 that interferon response to respiratory viruses highly upregulates *ACE2* expression.
42 Finally, we define airway responses to coronavirus infections in children, finding that
43 these infections upregulate *IL6* while also stimulating a more pronounced cytotoxic
44 immune response relative to other respiratory viruses. Our results reveal mechanisms
45 likely influencing SARS-CoV-2 infectivity and COVID-19 clinical outcomes.

46

47

48

49

50

51

52

53

54 *Introduction*

55

56 In December of 2019, a novel Coronavirus, SARS-CoV-2, emerged in China and has
57 gone on to trigger a global pandemic of Coronavirus Disease 2019 (COVID-19), the
58 respiratory illness caused by this virus¹. While most individuals with COVID-19
59 experience mild cold symptoms (cough and fever), some develop more severe disease
60 including pneumonia, which often necessitates mechanical ventilation². In fact, an
61 estimated 5.7% of COVID-19 illnesses are fatal³. Enhanced risk of poor outcomes for
62 COVID-19 has been associated with a number of factors including advanced age, male
63 sex, and underlying cardiovascular and respiratory conditions^{4, 5}. Yet, while the majority
64 of serious COVID-19 illness occurs in adults over 60, children are also thought to be
65 highly susceptible to infection. Moreover, recent data suggest that 38% of COVID-19
66 cases occurring in children are of moderate severity and 5.8% are severe or critical⁶,
67 highlighting a need for studying risk factors of illness in this population as well.

68

69 One factor that may underlie variation in clinical outcomes of COVID-19 is the extent of
70 gene expression in the airway of the SARS-CoV-2 entry receptor, ACE2, and
71 TMPRSS2, the host protease that cleaves the viral spike protein and thus allows for
72 efficient virus-receptor binding⁷. Expression of these genes and their associated
73 programs in the nasal airway epithelium is of particular interest given that the nasal
74 epithelium is the primary site of infection for upper airway respiratory viruses, including
75 coronaviruses, and acts as the gateway through which upper airway infections can
76 spread into the lung. The airway epithelium is composed of multiple resident cell types

77 (e.g., mucus secretory, ciliated, basal stem cells, and rare epithelial cell types)
78 interdigitated with immune cells (e.g. T cells, mast cells, macrophages), and the relative
79 abundance of these cell types in the epithelium can greatly influence the expression of
80 particular genes⁸⁻¹⁰, including *ACE2* and *TPRSS2*. Furthermore, since the airway
81 epithelium acts as a sentinel for the entire respiratory system, its cellular composition,
82 along with its transcriptional and functional characteristics, are significantly shaped by
83 interaction with environmental stimuli. These stimuli may be inhaled (e.g., cigarette
84 smoke, allergens, microorganisms) or endogenous, such as when signaling molecules
85 are produced by airway immune cells present during different disease states. One such
86 disease state is allergic airway inflammation caused by type 2 (T2) cytokines (IL-4, IL-5,
87 IL-13), which is common in both children and adults and has been associated with the
88 development of both asthma and COPD in a subgroup of patients¹¹⁻¹³. T2 cytokines are
89 known to greatly modify gene expression in the airway epithelium, both through
90 transcriptional changes within cells and epithelial remodeling in the form of mucus
91 metaplasia^{11, 14, 15}. Microbial infection is another strong regulator of airway epithelial
92 expression. In particular, respiratory viruses can modulate the expression of thousands
93 of genes within epithelial cells, while also recruiting and activating an assortment of
94 immune cells¹⁶⁻¹⁸. Even asymptomatic nasal carriage of respiratory viruses, which is
95 especially common in childhood, has been shown to be associated with both genome-
96 wide transcriptional re-programming and infiltration of macrophages and neutrophils in
97 the airway epithelium¹⁹, demonstrating how viral infection can drive pathology even
98 without overt signs of illness.

99

100 Genetic variation is another factor that may regulate gene expression in the airway
101 epithelium. Indeed, expression quantitative trait loci (eQTL) analyses carried out in
102 many tissues have suggested that as many as 70% of genes expressed by a tissue or
103 organ are under genetic control²⁰. Severity of human rhinovirus (HRV) respiratory illness
104 has specifically been associated with genetic variation in the epithelial genes *CDHR3*²¹
105 and the *ORMDL3*²² and, given differences in genetic variation across world populations,
106 it is possible that functional genetic variants in SARS-CoV-2-related genes could partly
107 explain population differences in COVID-19 clinical outcomes.

108

109 Finally, there are important questions regarding the host response to SARS-CoV-2
110 infection. For example, it is unclear whether specific antiviral defenses in the epithelium
111 are blocked by SARS-CoV-2 or whether the virus may trigger epithelial or immune cell
112 pathways that prolong airway infection, and/or even incite a hyperinflammatory state in
113 the lungs in some individuals that leads to more severe disease. Although large cohorts
114 of subjects infected by the novel coronavirus are still lacking, much can be learned by
115 exploring transcriptional responses to other coronavirus strains. In particular, because
116 nasal airway brushings capture both epithelial and immune cells present at the airway
117 surface, such samples collected from a cohort of subjects infected by a range of viruses
118 provide an opportunity to comprehensively investigate the potentially varied and
119 cascading effects of coronavirus infection on airway expression and function.

120

121 In this study, we first use single cell RNA-sequencing (scRNA-seq) to elucidate the
122 cellular distribution of *ACE2* and *TMPRSS2* expression in the nasal airway epithelium.

123 We also perform network and eQTL analysis of bulk gene expression data on nasal
124 airway epithelial brushings collected from a large cohort of asthmatic and healthy
125 children in order to identify the genetic and biological regulatory mechanisms governing
126 *ACE2* and *TMPrSS2* expression. We then use multi-variable modeling to estimate the
127 relative contribution of these factors to population variation in the expression of these
128 two genes, and by performing experiments on mucociliary airway epithelial cultures
129 confirm a dominant role for both T2 inflammation and viral infection in regulating
130 expression of *ACE2* and *TMPrSS2*. Finally, we define the cellular and transcriptional
131 responses to *in vivo* coronavirus infections in the nasal airway of children.

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146 *Results*

147

148 **ACE2 and TMPRSS2 are expressed by multiple nasal airway epithelial cell types**

149

150 We first examined *ACE2* and *TMPRSS2* expression at a cell type level through single
151 cell RNA sequencing (scRNA-seq) of a nasal airway epithelial brushing from an
152 asthmatic subject. Shared Nearest Neighbor (SNN)-based clustering of 8,291 cells
153 identified 9 epithelial and 3 immune cell populations (Figure 1a, Supplementary Table
154 1). We found that 7 epithelial cell populations contained *ACE2*⁺ cells (at low frequency),
155 with the highest frequency of positive cells found among basal/early secretory cells,
156 ciliated cells, and secretory cells (Figure 1b). We did not observe meaningful *ACE2*
157 expression among any of the immune cell populations, which included T cells, dendritic
158 cells, and mast cells. We found *TMPRSS2* to be expressed by all epithelial cell types,
159 with a higher frequency of positive cells among the different cell types, compared to
160 *ACE2* (Figure 1b,c). A small number of mast cells were also *TMPRSS2*⁺ (Figure 1c).

161

162

163 ***TMPRSS2* is part of a mucus secretory co-expression network highly induced by
164 T2 inflammation**

165

166 We next sought to determine the variation in nasal epithelial expression of *ACE2* and
167 *TMPRSS2* across healthy and asthmatic children, and to identify biological mechanisms
168 that regulate this variation. Thus, we performed weighted gene co-expression network

169 analysis (WGCNA) on whole transcriptome sequencing data from nasal airway
170 brushings of 695 Puerto Rican healthy and asthmatic children in the Genes-
171 Environments and Asthma in Latino Americans II study (GALA II). This analysis
172 identified 54 co-expression networks representing cell type-specific expression
173 programs such as ciliogenesis, mucus secretion, and pathways of immunity and airway
174 inflammation (Supplementary Table 2). The *TPRSS2* gene was contained within one
175 of a set of three highly correlated networks exhibiting strong enrichments for mucus
176 secretory cell genes and pathways (Figure 2a, Supplementary Table 2,3). For example,
177 the black network, which was highly correlated with *TPRSS2* expression ($r=0.64$,
178 $p=1e-82$), was strongly enriched for *Golgi mediated transport* and *COPI-dependent*
179 *Golgi to ER transport* pathways, both of which are involved in the normal processing
180 and transport of mucin proteins (Figure 2a). *TPRSS2* itself fell within and was highly
181 correlated with expression of the pink network ($r=0.68$, $p=3e-97$), which was highly
182 enriched for *mucus goblet cell* markers ($p=2e-6$, Figure 2a,b). The pink network was
183 also enriched for genes involved in the *O-linked glycosylation of mucins* pathway ($p=9e-4$),
184 which is vital to the function of mucus secretory cells, especially those induced by T2
185 inflammation ($r=0.68$, $p=3e-97$, Figure 2a,b). In fact, we found that this network
186 contained the T2 cytokine *IL13* while being particularly enriched for genes known to
187 mark and transcriptionally regulate IL-13-induced mucus metaplasia (*FCGBP*, *SPDEF*,
188 *FOXA3*). The saddle brown network was also related to mucus secretory cells, and
189 contained the most canonical T2 inflammation markers^{11, 23} including *POSTN*, *CLCA1*,
190 *CPA3*, *IL1RL1*, *CCL26*, and was strongly correlated with both *TPRSS2* ($r=0.61$, $p=5e-72$,
191 Figure 2c) and the other T2 mucus secretory network (pink) ($r=0.92$, $p=3e-280$,

192 Supplementary Table 4). In contrast, we found *ACE2* expression to be strongly
193 negatively correlated with expression of both T2 networks (pink: $r=-0.61$, $p=3e-72$,
194 saddle brown: $r=-0.7$, $p=2e-102$, Figure 2e,f). To identify subjects with high and low T2
195 inflammation, we hierarchically clustered all subjects based on the expression of genes
196 in the canonical T2 network (saddle brown). This resulted in the identification of two
197 distinct groups we labeled as T2-high ($n=364$) and T2-low ($n=331$) (Supplementary
198 Figure 1a). We found that this expression-derived T2 status was strongly associated
199 with traits known to be driven by T2 inflammation including IgE levels, exhaled nitric
200 oxide (FeNO), blood eosinophils, and asthma diagnosis (Supplementary Figure 1b-e).
201 Notably, *TMPRSS2* levels were 1.3-fold higher in T2-high subjects ($p=1e-62$), while,
202 *ACE2* expression was 1.4-fold lower in T2-high subjects ($p=2e-48$) (Figure 2d,g).

203
204 To investigate whether the strong *in vivo* relationship between airway T2 inflammation
205 and *TMPRSS2/ACE2* expression is causal in nature, we performed *in vitro* stimulation
206 of paired air-liquid interface (ALI) mucociliary airway epithelial cultures with 72 hours of
207 IL-13 or mock stimulus ($n=5$ donors, Figure 3a). Performing paired differential
208 expression analysis between the mock and IL-13 stimulated cultures, we found that
209 *ACE2* and *TMPRSS2* were strongly down- and up-regulated, respectively, supporting
210 our *in vivo* analysis results ($\log_2\text{FC} = -0.67$, $p=5e-3$, $\log_2\text{FC} = 1.20$, $p=5e-9$, Figure 3b,c).
211 To better understand the cellular basis of *TMPRSS2* and *ACE2* regulation by IL-13, we
212 leveraged scRNA-seq data previously generated on tracheal airway epithelial cultures
213 that were chronically stimulated (10 days) with IL-13 or control media (Figure 3a,d).
214 Similar to our results from *in vivo* nasal scRNA-seq data, we observed that *ACE2*

215 expression was highest among basal, ciliated, and early/intermediate secretory cell
216 populations, with *ACE2* being significantly downregulated by IL-13 among both basal
217 and intermediate secretory cells (Figure 3e). Also mirroring the *in vivo* scRNA-seq data,
218 *TMPRSS2* was expressed across all epithelial cell types, but at a higher frequency
219 among secretory cells (Figure 3f). IL-13 stimulation induced dramatic upregulation of
220 *TMPRSS2* in early secretory, intermediate secretory, and mature mucus secretory cell
221 populations (Figure 3f). Furthermore, IL-13 stimulated mucus metaplasia that resulted in
222 the development of a novel mucus secretory cell type and an IL-13 inflammatory
223 epithelial cell that both highly expressed *TMPRSS2* (Figure 3f). Together, our *in vivo*
224 and *in vitro* analyses strongly suggest that *TMPRSS2* is part of a mucus secretory cell
225 network that is highly induced by IL-13-mediated T2 inflammation.

226
227

228

229

230 ***ACE2* belongs to an interferon response network that is induced by respiratory
231 virus infections**

232

233 Returning to the *in vivo* nasal airway epithelial expression networks, we found that
234 *ACE2* expression was highly correlated with expression of two networks (purple and
235 tan) (purple: $r=0.74$, $p=3e-120$, tan: $r=0.72$, $p=2e-110$, Figure 4a,b). The purple network
236 was highly enriched for genes that mark cytotoxic T cells and antigen-presenting
237 dendritic cells, both of which are particularly abundant in a virally infected epithelium
238 (Figure 4c, Supplementary Table 2), whereas the tan network was strongly enriched for

239 interferon and other epithelial viral response genes (*IFI6*, *IRF7*, *CXCL10*, *CXCL11*)
240 (Figure 4c, Supplementary Table 2). Clustering of subjects based on the interferon
241 response network genes resulted in two groups, one highly (interferon-high=78) and
242 one lowly (interferon-low=617) expressing these interferon response network genes
243 (Supplementary Figure 2). We found that *ACE2* expression was 1.7-fold higher in the
244 interferon-high vs. interferon-low group (Figure 4d). In a previous study, we found that
245 children with nasal gene expression characteristic of the interferon network tended to be
246 infected with a respiratory virus, despite being asymptomatic¹⁹. To explore the
247 possibility of this relationship in our current dataset, we metagenomically analyzed the
248 RNA-seq data for all subjects to identify those harboring reads for a respiratory virus.
249 This analysis found that 18% of GALA II children were asymptotically harboring a
250 respiratory virus from one of eight general respiratory virus groups (Figure 4e).
251 Strikingly, we found that 78% of interferon-high subjects were virus carriers compared to
252 only 10% of interferon-low subjects. These results demonstrate how asymptomatic virus
253 carriage nonetheless stimulates an active viral response that includes *ACE2*.

254
255 To directly test the effect of respiratory virus infection on epithelial *ACE2* gene
256 expression we again employed our ALI mucociliary epithelial culture system. Performing
257 mock or human rhinovirus-A16 infection of mature cultures (Day 27, Figure 4f) from 5
258 donors we found 7.7-fold upregulation of *ACE2* gene expression with HRV-A infection
259 ($p=1.3e-51$, Figure 4g). In contrast, we only observed a trend for down regulation of
260 *TMPRSS2* gene expression among virally infected subjects (Figure 4h). These results
261 confirm the strong regulation of *ACE2* gene expression by viral infection.

262

263 **Genetic determinants of ACE2 and TMPRSS2 expression in the nasal airway**
264 **epithelium**

265

266 We next explored the role of genetic regulatory variants in helping to drive epithelial
267 expression of ACE2 and TMPRSS2. To do this, we performed cis-eQTL analysis for
268 these two genes, using nasal gene expression and genome-wide genetic variation data
269 collected from the GALA II study children. We identified 316 and 36 genetic variants
270 significantly associated with expression of ACE2 and TMPRSS2, respectively (Figure
271 5a,b). Stepwise forward-backward regression analysis of these eQTL variants revealed
272 a single independent eQTL variant (rs181603331) for the ACE2 gene (6e-23), located
273 ~20kb downstream of the transcription start site (Figure 5a). This rare eQTL variant
274 (allele frequency [AF]=1%) was associated with a large decrease in ACE2 expression
275 ($\log_2 A_{FC}=-1.6$) (Figure 5c).

276 Similar analysis of the TMPRSS2 eQTL variants yielded three independent eQTL
277 variants (rs1475908 AF=20%, rs74659079 AF=4%, and rs2838057 AF=13%, Figure
278 5b). The eQTL variant rs1475908 was associated with a decrease in TMPRSS2
279 expression ($\log_2 A_{FC}=-0.37$, Figure 5d), whereas both the rs74659079 and rs2838057
280 eQTL variants were associated with increased TMPRSS2 expression ($\log_2 A_{FC}=0.38$,
281 0.43, respectively, Supplementary Figure 3).

282

283 Examining the frequency of these eQTL variants among eight world populations listed in
284 the gnomAD genetic variation database (v2.1.1), we found that the ACE2 eQTL variant

285 was only present in people of African descent and at a low frequency (AF=0.7%, Figure
286 5e). In contrast, the *TMPRSS2* eQTL variant associated with decreased expression,
287 rs1475908, occurred across all world populations, with the highest allele frequencies
288 among East Asians (AF=38%), Europeans (AF=35%), intermediate frequencies among
289 Africans (AF=26%) and Ashkenazi Jews (AF=23%), and the lowest frequency among
290 Latinos (AF=17%). The two *TMPRSS2* eQTL variants associated with increased
291 expression exhibited much more disparate allele frequencies across world populations.
292 Namely, the allele frequency of rs74659079 is above 1% only among people of African
293 descent (AF=11%) and 4% in the participating Puerto Rican population. Likewise, the
294 rs2838057 eQTL variant, which was associated with increased *TMPRSS2* expression
295 was present at a frequency of 32% in East Asians, 20% in Latinos, and <10% in all
296 other world populations. Together, these results suggest that if *TMPRSS2* levels
297 influence susceptibility to SARS-CoV-2, then genetics may play a significant role in
298 infection risk and that this risk will vary significantly across world populations.

299

300

301 **Multi-variable modeling of airway *ACE2* and *TMPRSS2* gene expression**

302

303 Our analyses indicate that T2 inflammation, interferon/viral response signaling, and
304 genetics are all determinants of *ACE2* and *TMPRSS2* gene expression in the airway
305 epithelium of children. Therefore, we next sought to determine the relative importance of
306 these factors in determining levels of these genes using multi-variable regression
307 analysis. We included asthma status, age, and sex as model covariates since chronic

308 lung disease, increasing age, and male sex have all been associated with increased risk
309 of poor COVID-19 illness outcomes. Modeling *ACE2* expression among GALA II
310 children, we found that T2 and interferon statuses had the strongest effects on *ACE2*
311 expression ($p=1.6e-57$, $p=6.5e-43$, respectively), with T2-low and interferon-high
312 individuals exhibiting the highest levels of expression. These two variables
313 independently explained 24% and 17% of the variance in *ACE2* expression (Table 1).
314 While the *ACE2* eQTL variant, rs181603331, was associated with a notable decrease in
315 *ACE2* levels, it only accounted for 1.2% of the variance, reflecting the low frequency of
316 this variant in our population. Increasing age and asthma diagnosis were both
317 associated with small decreases in *ACE2* expression, although both variables
318 accounted for less than 2% of the variance, and sex was not a significant predictor
319 (Table 1).

320

321 Similar modeling of *TPRSS2* expression found that T2-high status dramatically
322 increased expression, with an effect size 5.4x larger than any other variable, capturing
323 33% of total variation in *TPRSS2* (Table 1). While statistically significant, the two
324 *TPRSS2* eQTL variants associated with increased expression exhibited small effect
325 sizes totaling <1% of variance explained. All other predictors were not significant.
326 Collectively, these modeling results confirm that both T2 and interferon inflammation are
327 strong and antagonistic regulators of *ACE2* expression and show that T2 inflammation
328 is the lone dominant driver of airway expression of *TPRSS2*.

329

330 **Coronavirus Infections drive an enhanced cytotoxic immune response**

331

332 Our metagenomic analysis of RNA-seq data from the nasal brushings identified 18

333 children with viral sequence reads from one of four different coronavirus (CoV) species

334 (OC43, JKU1, 229E, NL63) (Supplementary Table 5). This allowed us to explore airway

335 transcriptomic responses to infection with coronavirus subfamily viruses specifically,

336 which will likely most resemble responses to SARS-CoV-2. To increase the likelihood

337 that these subjects were experiencing an active viral infection, we limited our analysis to

338 the 11 most highly infected subjects, comparing them to all subjects not infected with a

339 virus (n=571). To allow us to discriminate CoV-enhanced responses from those that are

340 more general to respiratory viruses, we also established a virus control group composed

341 of the 37 subjects highly infected with human rhinovirus species (HRV) (Supplementary

342 Table 6). We first compared expression of genes in the cytotoxic immune response

343 (purple) network and interferon response (tan) network (discussed earlier; see Figure

344 4a, b) among these virus infected groups, and found that both networks were more

345 highly expressed in virus-infected individuals (Figure 6a, b). Moreover, while the

346 induction in interferon response was similar for both CoV and HRV groups, induction in

347 the cytotoxic immune response was considerably higher in CoV-infected ($\Delta E_g = 0.049$)

348 compared to HRV-infected individuals ($\Delta E_g = 0.032$, Figure 6b).

349 To further explore this increase in cytotoxic immune response and other potential

350 pathways in CoV-infected individuals, we next performed a transcriptome-wide screen

351 for genes differentially expressed in CoV or HRV-infected groups compared to

352 uninfected individuals. These analyses revealed 2,515 differentially expressed genes

353 (DEGs) with CoV infection and 2,357 DEGs with HRV infection (FDR < 0.05 and $\log_2 FC$

354 > |0.5|), of which 35% and 31% were only observed with CoV and HRV infections,
355 respectively, based on our significance cutoff (Figure 6c). Upstream regulator analysis
356 with IPA carried out separately on CoV and HRV infection response genes showed that
357 the top cytokines and transcription factors that may regulate these infections are shared
358 between the two virus families, including IL10, IL1B, IFNG, IFNA2, and STAT1 (Figure
359 6d). One inferred upstream regulator of CoV response genes, IL-6, which was also
360 among the genes upregulated with CoV infection ($\log_2\text{FC}=2.2$, Figure 6e), is especially
361 noteworthy considering that an IL-6 blocking antibody therapy is currently under
362 investigation for use in treatment of COVID-19 illnesses²⁴. Additionally, we found ACE2
363 among the shared upregulated genes, reinforcing its upregulation in the course of
364 different respiratory virus infections ($\log_2\text{FC}$ in $\text{CoV}^+ = 0.6$, $\log_2\text{FC}$ in $\text{HRV}^+ = 0.5$, Figure
365 6e).

366

367 In trying to understand the biological basis of the viral responses we found to be CoV-
368 specific in our differential expression analysis, we considered whether the differential
369 presence and/or response of various immune cell types was an explanatory factor. To
370 investigate this, we used gene set enrichment analysis (GSEA) to test for enrichment of
371 CoV-specific, HRV-specific, and CoV/HRV-shared DEG sets among gene markers for
372 11 different flow-sorted human immune cell types defined based on whole transcriptome
373 data (citation) (Supplementary Table 7). The shared viral DEGs showed significant
374 enrichment for genes characteristic of macrophages, monocytes, neutrophils, dendritic
375 cells, and NK cells. In contrast, the set of CoV-enhanced DEGs resulted in strong
376 enrichments for both CD8+ T cells and dendritic cells, suggesting an especially

377 important role for activation of cytotoxic T cells through antigen presentation by dendritic
378 cells in CoV infections (Figure 6f). Also supporting an enriched cytotoxic response
379 among CoV-infected subjects was a strong enrichment for CoV-specific DEGs among
380 NK cells, which participate heavily in the killing of virally infected cells (Figure 6g). We
381 note that these enrichments were not observed among HRV-enhanced DEGs, which
382 were instead most strongly enriched among neutrophils, as well as eosinophils,
383 macrophages, and monocytes. Furthermore, through pathway analysis we identified
384 multiple pathways related to cytotoxic T cell and NK cell activity that were enriched
385 either specifically or more dramatically among CoV DEGs compared to HRV DEGs
386 (Figure 6e). These results suggest that while CoV infections are highly similar to HRV
387 infections, they likely elicit an enhanced cytotoxic immune response.

388

389 *Discussion*

390

391 Although the high variability in clinical outcomes of COVID-19 illness is now well
392 documented and multiple demographic and clinical traits have been associated with
393 severe disease, little is known about the host biologic factors underlying this variability.
394 In the current study, we reasoned that population variation in upper airway expression
395 of the ACE2 receptor for SARS-CoV-2 and the virus-activating TMPRSS2 protease,
396 would drive infection susceptibility and disease severity. We therefore deployed network
397 and eQTL analysis of nasal airway epithelial transcriptome data from a large cohort of
398 healthy and asthmatic children to determine mechanisms associated with airway
399 expression of these genes, and their relative power in explaining variation in the

400 expression of these genes among children. We observed only weak associations with
401 asthma status, age, and gender among children aged 8-21 years. Moreover, although
402 we found that genetics does influence expression of these genes, the effect of this
403 variation was small in comparison to the dramatic influence of T2 cytokine-driven
404 inflammation on both *ACE2* (downregulation) and *TMPrSS2* (upregulation) expression
405 levels. We found an equally important role for viral-driven interferon inflammation in
406 regulating levels of *ACE2* in the airway. Additionally, through study of *in vivo* upper
407 airway CoV subfamily infections, we not only identify inflammatory regulators of these
408 infections, but also provide evidence that this subfamily of viruses drives an enhanced
409 cytotoxic immune response. Our work provides a set of biomarkers that can be easily
410 examined in COVID-19 patients, through analysis of nasal swabs, to determine the
411 relative importance of these mechanisms and genes in governing susceptibility to
412 infection, severe illness and death.

413

414 Our single cell analysis of an *in vivo* nasal brushing observed *ACE2* expression, albeit
415 at low frequency, primarily among basal, ciliated, and less mature, early secretory cells.
416 These results are supported by a recent report of *ACE2* expression in transient
417 secretory cells, likely a close equivalent to our early secretory population²⁵. Although a
418 much higher portion of cells, representing all epithelial cell types, expressed *TMPrSS2*,
419 the low frequency of *ACE2*⁺ cells resulted in very few dual *ACE2*/*TMPrSS2* expressing
420 cells. However, we caution that a cell may not need to be *TMPrSS2*⁺ to be susceptible
421 to infection, since it has been demonstrated the *TMPrSS2* protein is secreted from
422 nasal airway epithelial cells²⁶. We also caution that scRNA-seq data are known to

423 exhibit biases in gene detection, and thus the level and frequency of *ACE2* expression
424 across cells may be much higher than we observe here. In line with this possibility we
425 observe more moderate levels of *ACE2* expression in our bulk RNA-seq data on nasal
426 brushings.

427

428 Airway inflammation caused by type 2 cytokine production from infiltrating immune cells
429 plays a prominent role in the control of cellular composition, expression, and thus
430 biology of the airway epithelium^{11, 13, 23, 27}. Moreover, while T2 airway inflammation is an
431 important driver of T2-high asthma and COPD disease endotypes, it is also associated
432 with atopy in the absence of lung disease, a very common phenotype in both children
433 and adults. In fact, among the children in this study, we find that 43% of non-asthmatics
434 were scored as T2-high based on expression profile, further substantiating the high
435 prevalence of T2 airway inflammation outside of those with lung disease. Our data
436 suggest that airway epithelial *TMPRSS2* expression is highly upregulated by T2
437 inflammation, and specifically by IL-13. Both our network and single cell data show that
438 *TMPRSS2* is most prominent in less developed “early secretory” cells as well as in more
439 mature mucus secretory cells. Based on our *in vitro* data, IL-13 upregulates *TMPRSS2*
440 across nearly all types of epithelial cells, but the core of this effect appears to be in the
441 metaplastic mucus secretory cells that are generated as a consequence of IL-13
442 signaling^{14, 15}. In fact, our network data suggest that, although *TMPRSS2* expression is
443 highly correlated with that of a co-expressed network of mucus secretory genes
444 characterizing “normal”, non-metaplastic, mucus secretory cells, it’s correlated even
445 more strongly with a network that characterizes mucus secretory cells undergoing IL-13-

446 induced metaplasia. In contrast to enhanced levels of *TPMRSS2*, T2 inflammation,
447 whether observed *in vivo* or induced with IL-13 stimulation, precipitated a dramatic
448 reduction in levels of epithelial *ACE2*, thus complicating expectations for how T2
449 inflammation might affect overall risk for a poor COVID-19 outcome. Germane to this
450 question, a recent study of 85 fatal COVID-19 subjects found that 81.2% of them
451 exhibited very low levels of blood eosinophil levels⁴. Blood eosinophil levels are a
452 strong, well-known predictor of airway T2 inflammation and were strongly correlated
453 with T2 status in our study as well^{11, 23}. Together, these studies provisionally suggest
454 that T2 inflammation may predispose individuals to experience better COVID-19
455 outcomes through a decrease in airway levels of *ACE2* that override any countervailing
456 effect from increased expression of *TPMRSS2*. However, both *in vitro* experiments
457 examining IL-13 effects on SARS-CoV-2 infection and empirical data on COVID-19
458 outcomes among T2-high and T2-low patients will certainly be needed to determine
459 whether this common airway inflammatory endotype ultimately protects against or
460 exacerbates COVID-19 illness. As mentioned above, we note that measurement of
461 blood eosinophil levels could be used as an informative and more accessible (albeit less
462 powerful) proxy for investigating the association between airway T2 inflammation and
463 outcomes of COVID-19. Moreover, given the higher frequency of T2 inflammation
464 among asthmatic subjects, this population should be monitored especially closely given
465 the enhanced risk of complications due to respiratory virus infection in those with
466 asthma.

467

468 In addition to a strong negative influence of T2 inflammation on *ACE2* expression in the
469 airway, we found an equally strong positive influence of respiratory virus infections on
470 levels of this gene. Network analysis placed *ACE2* within an interferon viral response
471 network suggesting that these cytokines are a driving force behind *ACE2* upregulation.
472 This information is interesting in several regards. First, it suggests that SARS-CoV-2
473 and other coronaviruses using *ACE2* as a receptor could leverage the host anti-viral
474 response to increase the infectability of airway cells. Secondly, as data here and
475 elsewhere show, asymptomatic carriage of respiratory viruses is common, especially in
476 young children^{19, 28-31}. Children in the GALA II cohort included in this study ranged in
477 age from 8-21 years; among them we found 18% who were carrying respiratory viruses
478 without illness. However, as we show in this and our previous study¹⁹, even
479 asymptomatic carriage of respiratory viruses exerts a fundamental change in airway
480 epithelial expression and immune cell presence, including upregulation of *ACE2*
481 expression. In determining outcomes, this potential detrimental influence of virus
482 carriage may also be weighed against a potentially beneficial influence of virus carriage
483 through a more potent cross serologic-immune defense in these individuals, especially if
484 the virus carried is a coronavirus family member. Ultimately, the effect of current or
485 recent virus carriage on COVID-19 outcomes will need to be determined by *in vivo*
486 studies in patients, followed up with controlled *in vitro* studies of virally infected cells. At
487 any rate, the apparent dependence of *ACE2* expression on interferon signaling
488 suggests that targeted blockade of this interferon effect could control SARS-CoV-2
489 infection.

490

491 Our evaluation of genetic influences on airway *ACE2* and *TMPRSS2* expression
492 revealed a single rare eQTL for *ACE2* and several more frequent eQTL variants for
493 *TMPRSS2*. While both the effect size and explanatory power of these variants paled in
494 comparison to the influence of T2 inflammation and interferon signaling in multi-variable
495 modeling of expression for these genes, the effect of these variants may still be strong
496 enough to alter infection rates and or illness severity, especially in the populations
497 where these variants are most frequent. Thus, future genetic studies of COVID-19
498 should pay particular attention to these eQTL variants.

499

500 A particularly vexing question regards the mechanisms that underlie the unusual
501 severity of illness associated with SARS-CoV-2, especially when compared to most
502 circulating respiratory viruses. Clearly, severe disease often entails development of
503 pneumonia, possibly resulting from an expanded tropism of SARS-CoV-2 to include
504 lower lung airway and alveolar cells. The most severe patients also appear to
505 experience an exuberant immune response, characterized a “cytokine storm”²⁴,
506 occurring with and possibly driving the development of acute respiratory distress
507 syndrome (ARDS). Supposing that aspects of epithelial response to coronavirus family
508 members would be shared, including with SARS-CoV-2, we examined *in vivo*
509 coronavirus infection among the GALA II children. We found that CoV infections elicit a
510 broad airway transcriptome response, similar to HRV infections, and we identified a
511 panel of cytokines and transcription factors that likely regulate these responses. In
512 particular, we found that IL-6 was predicted to regulate responses to CoV and was itself
513 upregulated with these infections. These data support the recent investigation of

514 tocilizumab (IL-6R blocking antibody) for the treatment of COVID-19 illnesses²⁴.
515 Strikingly our analysis revealed an increased cytotoxic immune response with CoV
516 infection, driven by the differential presence and activity of cytotoxic CD8+ T cells and
517 NK cells, as compared to the more heavily neutrophil-based responses to HRV
518 infection. Although preliminary, this finding, if similarly occurring with SARS-CoV-2
519 infection, could partly explain the dramatic inflammation observed in SARS-CoV-2
520 patients, which can extend to the distal lung.

521

522 In summary, our data suggests that the strongest determinants of airway ACE2 and
523 *TMRSS2* expression are T2 inflammation and viral-induced interferon inflammation,
524 with limited but noteworthy influence from genetic variation. Whether these factors drive
525 better or worse clinical outcomes remains to be determined, but closely watching
526 individuals with these airway endotypes in the clinical management of COVID-19
527 illnesses would be prudent.

528

529

530

531

532

533

534

535

536

537 *Methods*

538

539 MATERIALS AND CORRESPONDENCE

540 Further information and requests for resources and reagents should be directed to and

541 will be fulfilled by Max A. Seibold, Ph.D. (seiboldm@njhealth.org)

542

543 EXPERIMENTAL METHODS

544

545 **Human subject information**

546 Under the Institutional Review Board (IRB) approved Asthma Characterization Protocol

547 (ACP) at National Jewish Health (HS-3101) we consented a 56 year old asthmatic

548 subject, from which we collected nasal airway epithelial cells. The nasal airway cells

549 were brushed from the inferior turbinate using a cytology brush and used for the scRNA-

550 seq experiment described in Figure 1. Nasal airway epithelial cells used for bulk RNA-

551 seq network and eQTL analysis came from GALA II study subjects described below.

552 Nasal airway epithelial cell ALI culture experiments all used cells derived from GALA II

553 study subjects. Human tracheal airway epithelial cells used for *in vitro* IL13 stimulation

554 and scRNA-seq experiment were isolated from a single de-identified lung donor

555 obtained from the International Institute for the Advancement of Medicine (Edison, NJ),

556 and Donor Alliance of Colorado. The National Jewish Health Institutional Review Board

557 (IRB) approved our research on the tracheal airway epithelial cells under IRB protocol

558 HS-3209. These cells were processed and given to us through the National Jewish

559 Health (NJH) live cell core, which is an institutional review board-approved study (HS-
560 2240) for the collection of tissues from consented patients for researchers at NJH.

561

562 **GALA II study subjects**

563 The Genes-Environment & Admixture in Latino Americans study (GALA II) is an on-
564 going case-control study of asthma in Latino children and adolescents. GALA II was
565 approved by local institutional review boards (UCSF, IRB number 10-00889, Reference
566 number 153543, NJH HS-2627) and all subjects and legal guardians provided written
567 informed assent and written informed consent, respectively^{32, 33}. A full description of the
568 study design and recruitment has been previously described elsewhere³²⁻³⁴. Briefly, the
569 study includes subjects with asthma and healthy controls of Latino descent between the
570 ages of 8 and 21, recruited from the community centers and clinics in the mainland U.S.
571 and Puerto Rico (2006-present). Asthma case status was physician-diagnosed.
572 Recruited subjects completed in-person questionnaires detailing medical,
573 environmental, and demographic information. Physical measurements including
574 spirometry were obtained, and subjects provided a blood sample for DNA extraction and
575 later Whole Genome Sequencing. GALA subjects that were part of this analysis were all
576 recruited from Puerto Rico (n=695). A nasal airway inferior turbinate brushing was used
577 to collect airway epithelial cells from these subjects for whole transcriptome sequencing
578 (n=695). Network analyses were performed on all subjects with nasal brushing whole
579 transcriptome sequencing data (n=695) and eQTL analysis was performed on the
580 subset (n=681) with whole genome sequencing generated genotype data.

581

582

583 **Bulk RNA sequencing of GALA II and ALI Samples**

584 Total RNA was isolated from GALA II subject nasal airway epithelial brushings using the
585 AllPrep DNA/RNA Mini Kit (QIAGEN, Germantown, MD). Whole transcriptome libraries
586 were constructed using the KAPA Stranded mRNA-seq library kit (Roche Sequencing
587 and Life Science, Kapa Biosystems, Wilmington, MA) from 250ng of total input RNA
588 with the Beckman Coulter Biomek FX^P automation system (Beckman Coulter, Fullerton,
589 CA) according to the manufacturers protocol. Barcoded libraries were pooled and
590 sequenced using 125bp paired-end reads on the Illumina HiSeq 2500 system (Illumina,
591 San Diego, CA). Bulk RNA-seq data for the nasal and tracheal ALI cultures to measure
592 ACE2 and TMPRSS2 levels reported in Figures 3b,c and 4g,h, was generated with
593 KAPA Hyperprep Stranded mRNA-seq library kits (Roche Sequencing and Life Science,
594 Kapa Biosystems, Wilmington, MA) and sequenced with a Novaseq 6000 using 150bp
595 paired end reads.

596

597

598 **Whole genome sequencing of GALA II Samples**

599 Genomic DNA was extracted from whole blood obtained from GALA II study subjects
600 using the Wizard Genomic DNA Purification kits (Promega, Fitchburg, WI), and DNA
601 was quantified by fluorescent assay. DNA samples were sequenced as part of the
602 Trans \square Omics for Precision Medicine (TOPMed) whole genome sequencing (WGS)
603 program³⁵. WGS was performed at the New York Genome Center and the Northwest
604 Genomics Center on a HiSeqX system (Illumina, San Diego, CA) using a paired \square end

605 read length of 150 base pairs, to a minimum of 30X mean genome coverage. Details on
606 DNA sample handling, quality control, library construction, clustering and sequencing,
607 read processing, and sequence data quality control are described elsewhere³⁵. Variant
608 calls were obtained from TOPMed data freeze 8 variant call format files.

609

610

611 **Experiments using an air-liquid interface, mucociliary culture system**

612 Primary human basal airway epithelial cells were expanded and differentiated at air-
613 liquid interface (ALI) *in vitro* according to established protocols³⁶. Paired tracheal ALI
614 cultures were mock-treated or treated with 10 ng/mL IL-13 in media (20 μ L apical; 500
615 μ L basolateral) for the final 10 days of differentiation (ALI days 11-21) before harvest
616 and scRNA-seq analysis. In contrast, nasal ALI cultures used for bulk RNA-seq analysis
617 (N = 5 GALA II subjects) were either stimulated with IL-13 for 72h following completion
618 of mucociliary differentiation (25 days) or were infected with human rhinovirus strain
619 A16 for 4 h during the final 24 h of the 28 days of differentiation. Control cultures were
620 only treated with media.

621

622 **Preparation of ALI cultures for 10X scRNAseq**

623 Following stimulation experiments involving the tracheal airway epithelial ALI samples,
624 apical culture chambers were washed once with PBS and once with PBS supplemented
625 with dithiothreitol (DTT;10mM), followed by two PBS washes to remove residual DTT.
626 Cold active protease (CAP) solution (*Bacillus licheniformis* protease 2.5 μ g/mL, DNase
627 125 U/mL, and 0.5 mM EDTA in DPBS w/o Ca^{2+} Mg^{2+}) was added to apical culture

628 chamber and incubated on ice for 10 minutes with mixing every 2.5 minutes.
629 Dissociated cells in CAP solution were added to 500 μ L cold FBS, brought up to 5 mL
630 with cold PBS, and centrifuged at 225 x g and 4°C for 5 minutes. The cell pellet was
631 resuspended in 1 mL cold PBS+DTT, centrifuged at 225 x g and 4°C for 5 minutes, and
632 then washed twice with cold PBS. The final cell pellet was resuspended in PBS with
633 0.04% BSA for single cell gene expression profiling with the 10X Genomics system.
634 Sample capture, cDNA synthesis, and library preparation for 10d IL-13 ALI stimulations
635 was performed using protocols and reagents for 10X Genomics Chromium Single Cell
636 3' v3 kit. Single cell libraries were pooled for sequencing on an Illumina NovaSeq 6000.

637

638 **Nasal brush 10X scRNA-seq**

639 Nasal brush cells were dissociated from the brush using *Bacillus licheniformis* cold
640 active protease (10mg/ml), EDTA (0.5mM), and EGTA (0.5mM) at 4°C with vortex
641 mixing, followed by enzyme neutralization with FBS. Red blood cell lysis was
642 performed and cells were washed twice in 0.04% BSA/PBS. Cell concentration was
643 adjusted to 400 cells/ μ L for cell capture of ~8000 cells using the 10X Genomics
644 Chromium Next GEM Single Cell 3' reagent kit chemistry. Sample capture, cDNA
645 synthesis, and library preparation was performed following 10X Genomics Chromium
646 Next GEM Single Cell 3' v3 kit. The single cell library was sequenced on an Illumina
647 NovaSeq 6000.

648

649 **QUANTIFICATION AND STATISTICAL ANALYSIS**

650

651 **Nasal airway epithelium brushing bulk RNA-seq analysis**

652

653 *Preprocessing of RNA-seq data*

654 Raw sequencing reads were trimmed using `skewer`³⁷ (v0.2.2) with the following
655 parameter settings: `end-quality=15`, `mean-quality=25`, `min=30`. Trimmed reads were
656 then aligned to the human reference genome GRCh38 using `GSNAP`³⁸ (v20160501)
657 with the following parameter settings: `max-mismatches=0.05`, `indel-penalty=2`, `batch=3`,
658 `expand-offsets=0`, `use-sarray=0`, `merge-distant-same-chr`. Gene quantification was
659 performed with `htseq-count`³⁹ (v0.9.1) using iGenomes GRCh38 gene transcript model.
660 Variance stabilization transformation (VST) implemented in `DESeq2`⁴⁰ (v1.22.2) was
661 then carried out on the raw gene count matrix to create a variance stabilized gene
662 expression matrix suitable for downstream analyses.

663

664 *Weighted Gene Co-expression Network Analysis (WGCNA) on GALA II RNA-seq data*

665 To understand what biological mechanisms regulate the variation of nasal airway
666 epithelial gene expression, Weighted Gene Co-expression Network Analysis⁴¹
667 (WGCNA) v1.68 was performed on the VST matrix of 17,473 expressed genes.
668 WGCNA analysis is a network-based approach that assumes a scale-free network
669 topology. To adhere to the scale-free assumption of the constructed biological networks,
670 a soft thresholding parameter (β) value of 9 was chosen based on WGCNA guidelines.
671 Furthermore, `minClusterSize` was set to 20, `deepSplit` was set to 2, and `pamStage` was
672 set to TRUE. A total of 54 co-expression networks were identified and described in
673 Supplementary Table 2. WGCNA networks are referred to by different colors, and two of

674 the these identified networks, saddle brown and tan were found to capture co-
675 expressed genes that underlie T2 inflammation and interferon inflammation,
676 respectively. We hierarchically clustered all subjects based on expression of genes in
677 the saddle brown network and then used the first split in the dendrogram as the basis
678 for assigning individuals to T2-high or T2-low categories (Supplementary Figure 1a).
679 Similarly, we hierarchically clustered subjects using the genes in tan network and then
680 selected the dendrogram branches with the highest tan network expression as
681 interferon-high and the other subjects as interferon-low (Supplementary Figure 2a).

682

683 *Cis-eQTL analysis of nasal RNA-seq data*

684 Cis-expression quantitative trait locus (eQTL) analysis was performed by following the
685 general methodology of the Genotype-Tissue Expression (GTEx) project version 7
686 protocol⁴², using the nasal RNA-seq data and WGS variant data from 681 GALA II
687 subjects.

688 Namely, WGS variant data was filtered based on allele frequency (minor allele
689 frequency > 1%) and allele subject count (total number of subjects carrying minor allele
690 ≥ 10). After filtering, 12,590,800 genetic variants were carried forward into the eQTL
691 analysis. For expression data filtering and preparation, we first ran Kallisto⁴³ (v0.43.0) to
692 generate transcript per million (TPM) values. We filtered out any genes that did not
693 reach both TPM > 0.1 and raw counts > 6 for at least 20% of our samples. After filtering,
694 17,039 genes were then TMM normalized using edgeR⁴⁴ (v3.22.3). Finally, we applied
695 an inverse normal transformation into the TMM-normalized expression values to render
696 them suitable for eQTL analysis. To account for global population structure, we ran

697 ADMIXTURE⁴⁵ (v1.3.0) on the genotype data to create five admixture factors. We then
698 ran Probabilistic Estimation of Expression Residuals⁴⁶ (PEER, v1.3) to create 60 PEER
699 factors to utilize as covariates in the eQTL analysis along with admixture estimates,
700 gender, age, body-mass index (BMI), and asthma diagnosis status. To perform cis-
701 eQTL analysis, we utilized a modified version of FastQTL⁴⁷ that was provided by the
702 GTEx project. Furthermore, we performed stepwise regression analysis to identify
703 independent eQTL variants using QTLTools⁴⁸ (v1.1). Allelic Fold Change (A_{FC}) of the
704 eQTL variant is computed using the aFC python script⁴⁹.

705

706 *Virus identification and quantification from bulk RNA-Seq data*

707 To identify individuals with asymptomatic virus infection at the time of sample collection,
708 viral genomic sequences were recovered from bulk RNA-seq data using a modified
709 version of the Virus Finder 2.0 (VF2) pipeline⁵⁰. A custom respiratory virus reference
710 database comprising >600k sequences was employed to improve specificity. Using
711 VF2, viral reads were garnered by removing human reads using Bowtie2⁵¹ (default
712 settings) and selecting viral reads using BLAT⁵² (minIdentity=80); contigs were
713 assembled using Trinity⁵³; short (<200 bp) or low complexity (DUST score < 0.07)
714 contigs and contigs matching the human genome at a BLAST⁵⁴ e-value <0.05 were
715 discarded; the remaining contigs were classified using BLAST (e-value <0.05); read
716 counts were obtained by read mapping using BLAT (minIdentity=98). Of the 468 distinct
717 viral reference sequences detected by VF2, 7 were identified as erroneous and
718 removed. The remaining 461 matches were manually assigned viral serotypes and the
719 results aggregated with R.

720

721 *Defining CoV and HRV infected groups and associated analysis*

722 To ensure we selected subjects that were experiencing an active host response to a
723 CoV infection, we examined the distribution of viral reads for the 18 CoV⁺ infected
724 subjects. We observed a clear break between the 7 subjects with the lowest viral read
725 counts (<3,000 reads) and 11 subjects with the highest viral read counts (>60,000
726 reads). Therefore, we selected these 11 highly infected subjects for analysis of host
727 responses to CoV infection. To generate a similar infection-control group, composed of
728 subjects highly infected with a different virus species, we examined the 67 HRV infected
729 subjects in GALA, enforcing a comparable lower bound of viral reads as with CoV,
730 adjusting for the smaller HRV genome size. Specifically, HRV genomes are ~7,000
731 base pairs, whereas CoV genomes are ~30,000 base pairs, making the HRV genome
732 ~25% of the size of the CoV genome. Therefore, we selected a cutoff of 15,000 viral
733 reads for subjects to be included in the HRV⁺ highly infected group. Therefore, we
734 selected a cutoff of 15,000 viral reads for subjects to be included in the HRV⁺ highly
735 infected group (n=37) analyzed in Figure 6. All non-infected subjects (n=571), based on
736 the Virus Finder analysis described above, were used as comparison group for the
737 CoV⁺ and HRV⁺ groups.

738 In performing the CoV⁺ and HRV⁺ transcriptome-wide differential expression analyses,
739 to account for the class imbalance of this experiment, log₂ count-normalized expression
740 values in units of counts per million (calculated using edgeR v3.28.0) were passed to
741 the function arrayWeights function in the limma⁵⁵ R package (3.42.0). limma-voom was
742 then used to perform differential expression analysis on the count normalized

743 expression values between the CoV⁺ and uninfected groups, as well as between the
744 HRV⁺ and uninfected groups, controlling for age, gender, and asthma diagnosis status.
745 Genes were required to have an FDR adjusted p-value < 0.05, and an absolute log₂FC
746 > 0.5 to be considered significant. Based on these cutoffs, genes were classified as
747 being shared if they were significant in both comparisons, or as CoV⁺-specific or HRV⁺-
748 specific if significant in only one comparison.

749

750 *Gene set enrichment analysis.*

751 To investigate enriched pathways within WGCNA networks (see Figure 2a) or within
752 genes differentially expressed in CoV⁺ and/or HRV⁺ infected subject groups (see Figure
753 6c and 6c), we used Enrichr⁵⁶ to test for gene overrepresentation of network genes
754 within a panel of annotated gene databases (Gene Ontology [GO] Biological Process
755 [BP] 2018, GO Molecular Function [MF] 2018, GO Cellular Component [CC] 2018,
756 Kyoto Encyclopedia of Genes and Genomes [KEGG] 2019 Human, and Reactome
757 2016). For cell type enrichments within WGCNA networks reported in Figure 2a, we
758 tested for overrepresentation of network genes within gene marker sets (FDR < 0.05)
759 for each of 35 epithelial and immune cell types inferred using scRNA-seq of human lung
760 tissue⁵⁷.

761

762 For the plots in Figure 6f-g, transcriptomic data for 11 flow sorted immune cell
763 populations were obtained from GEO experiments GSE3982 and GSE22886 and then
764 batch corrected using the ComBat⁵⁸ function from the SVA R package (v3.34.0). limma
765 was then used to perform differential expression analysis between each cell type and all

766 the rest in order to obtain gene \log_2FC values for each cell type with which to rank order
767 the genes. Gene set enrichment analysis (GSEA) was then used to test for association
768 between upregulated genes in the shared, CoV⁺-specific, and HRV⁺-specific gene sets
769 and each of the cell types, based on the cell type-specific ordered gene lists. GSEA was
770 carried out using the FGSEA R package (v1.12.0).

771

772

773 *Canonical pathway analysis.*

774 We used QIAGEN's Ingenuity Pathway Analysis (IPA) program (v01-16; content
775 version: 51963813, release 2020-03-11) to investigate canonical pathways and
776 upstream regulators that were significantly enriched in one or both of the upregulated
777 CoV⁺-specific or HRV⁺-specific gene sets.

778

779 **Analysis of scRNA-seq data from the nasal epithelial brushing**

780 Initial processing of 10X scRNA-seq data, including cell demultiplexing, alignment to the
781 human genome GRCh38, and UMI-based quantification was performed with Cell
782 Ranger (version 3.0). Since the nasal brushing sample contains both epithelial and
783 immune cell populations that have distinct expression profiles (e.g.: Immune cell types
784 express far fewer genes compared to epithelial cell types), clustering and cell type
785 identification were done in two stages: 1) an initial clustering with a less stringent filter to
786 identify major epithelial and immune cell clusters was performed, 2) cells were
787 reclustered with different independent filtering criteria for epithelial and immune cell
788 types. All these analyses were performed using Seurat⁵⁹ R package (v3.0).

789

790 In the first stage, we removed cells with fewer than 100 genes detected or cells with
791 greater than 25% mitochondrial reads. Additionally, to remove possible doublets, we
792 removed cells with higher than 6,000 genes detected and/or more than 20,000 UMIs.
793 Lowly expressed genes (detected in fewer than 4 cells) were also removed. We then
794 performed normalization using SCTtransform⁶⁰ and ran PCA on the top 5000 highly
795 variable normalized genes. Clustering analysis was performed on the top 20 PCs using
796 a shared nearest neighbor (SNN) based SLM⁶¹ algorithm with the following parameter
797 settings: resolution=0.8, algorithm=3. The single cell expression profiles were visualized
798 via embedding into two dimensions with UMAP⁶² (Uniform Manifold Approximation and
799 Projection), resulting in the identification of 11,157 epithelial cells and 229 immune cells
800 based on known cell type signatures.

801

802 In the second stage, we retained all the immune cells but removed epithelial cells with
803 fewer than 1,000 detected genes. After this filtering, a combined 8,291 epithelial and
804 immune cells were then normalized as in the first stage. Clustering analysis performed
805 on the top 30 PCs with parameters (resolution=0.4, algorithm=1, k.param=10) identified
806 15 clusters. We then ran differential expression analysis using a Wilcoxon test
807 implemented in Seurat's "FindMarkers" function to help with cell type identification.
808 Based on these cluster marker lists, two clusters were merged into a single secretory
809 cluster, another two clusters were merged into a single ciliated cluster, and a final two
810 clusters were combined as "indeterminate," based on the lack of defining marker genes
811 for these clusters. Through this merging process, we arrived at 8 epithelial and 3

812 immune cell populations (Figure 1a, Supplementary Table 1)

813

814

815 **Analysis of bulk RNA-seq data from IL-13 and HRV infected ALI nasal airway**
816 **epithelial cultures**

817 Raw sequencing reads were trimmed using skewer with the following parameter
818 settings: end-quality=15, mean-quality=25, min=30. Trimmed reads were then aligned to
819 the human reference genome GRCh38 using HISAT2⁶³ (v2.1.0) using default parameter
820 settings. Gene quantification was performed with htseq-count using the GRCh38
821 Ensembl v84 gene transcript model. After removing mitochondrial, ribosomal, and lowly
822 expressed genes (those not expressed in at least two samples), we carried out
823 differential expression analyses between paired IL-13-stimulated and control samples
824 (N = 5 donors) and between paired HRV-infected and control samples (N = 5 donors)
825 using the DESeq2 R package (v1.22.2).

826

827 **Analysis of scRNA-seq data from 10 day IL-13-stimulated and control tracheal cell**
828 **ALI cultures**

829 As with the nasal brushing scRNA-seq data, 10X scRNA-seq data from ALI cultures
830 grown from a single tracheal donor that were either mock- or IL-13 stimulated for 10
831 days were pre-processed using Cell Ranger (version 3.0, 10X Genomics). To safeguard
832 against doublets, we removed all cells with gene or UMI counts exceeding the 99th
833 percentile. We also removed cells expressing fewer than 1,500 genes or for which >
834 30% of genes were mitochondrial (genes beginning with *MTAT*, *MT-*, *MTCO*, *MTCY*,

835 *MTERF, MTND, MTRF, MTRN, MRPL, or MRPS*), resulting in a total of 6,969 cells
836 (2,715 IL-13-stimulated and 4,254 controls). After removing mitochondrial, ribosomal
837 (*RPL* and *RPS*), and very lowly expressed genes (expressed in < 0.1% of cells), we
838 integrated expression data from IL-13 and control cells using the dataset integration
839 approach in Seurat⁶⁴. For the integration analysis, we used the top 30 dimensions from
840 a canonical correlation analysis (CCA) based on SCTtransform normalized expression of
841 the top 3,000 most informative genes across the two datasets, where “informativeness”
842 was defined by gene dispersion (i.e., the log of the ratio of expression variance to its
843 mean) across cells, calculated after accounting for its relationship with mean
844 expression. We then carried out principle component analysis (PCA) on the integrated
845 dataset and used the top 20 components for clustering and visualization. We used SNN
846 (Louvain algorithm, resolution=0.23, k.param=10) to cluster the integrated cells into 11
847 populations, which we visualized in two dimensions using UMAP (see Figure 3d). These
848 clusters were assigned cell type labels based their most upregulated genes, which were
849 identified by carrying out differential expression analysis between each cluster and all
850 others using Seurat’s logistic regression (LR) test, in which cell treatment was included
851 as a latent variable.

852

853 DATA AVAILABILITY

854 All raw and processed RNA-seq data used in this study are currently being deposited in
855 the National Center for Biotechnology Information/Gene Expression Omnibus (GEO).

856

857 CODE AVAILABILITY

858

859

860

861

862

863

864 *References*

865

- 866 1. Wang, C., Horby, P.W., Hayden, F.G. & Gao, G.F. A novel coronavirus outbreak
867 of global health concern. *Lancet* **395**, 470-473 (2020).
- 868 2. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019.
869 *N Engl J Med* **382**, 727-733 (2020).
- 870 3. Baud, D. *et al.* Real estimates of mortality following COVID-19 infection. *Lancet*
871 *Infect Dis* (2020).
- 872 4. Du, Y. *et al.* Clinical Features of 85 Fatal Cases of COVID-19 from Wuhan: A
873 Retrospective Observational Study. *Am J Respir Crit Care Med* (2020).
- 874 5. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with
875 COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054-1062
876 (2020).
- 877 6. Dong, Y. *et al.* Epidemiological Characteristics of 2143 Pediatric Patients With
878 2019 Coronavirus Disease in China. *Pediatrics* (2020).
- 879 7. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2
880 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* (2020).

881 8. Goldfarbmuren, K.C. *et al.* Dissecting the cellular specificity of smoking effects
882 and reconstructing lineages in the human airway epithelium. *Nature Communications* **in press** (2020).

883

884 9. Montoro, D.T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319-324 (2018).

885

886 10. Plasschaert, L.W. *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377-381 (2018).

887

888 11. Woodruff, P.G. *et al.* T-helper Type 2-driven inflammation defines major subphenotypes of asthma. *American Journal of Respiratory and Critical Care Medicine* **180**, 388-395 (2009).

889

890

891 12. Barcelo, B. *et al.* Intracellular cytokine profile of T lymphocytes in patients with chronic obstructive pulmonary disease. *Clinical and Experimental Immunology* **145**, 474-479 (2006).

892

893

894 13. George, L. & Brightling, C.E. Eosinophilic airway inflammation: role in asthma and chronic obstructive pulmonary disease. *Therapeutic Advances in Chronic Disease* **7**, 34-51 (2016).

895

896

897 14. Chen, G. *et al.* SPDEF is required for mouse pulmonary goblet cell differentiation and regulates a network of genes associated with mucus production. *Journal of Clinical Investigation* **119**, 2914-2924 (2009).

898

899

900 15. Lachowicz-Scroggins, M.E. *et al.* Abnormalities in MUC5AC and MUC5B protein in airway mucus in asthma. *American Journal of Respiratory and Critical Care Medicine* **194**, 1296-1299 (2016).

901

902

903 16. de Lamballerie, C.N. *et al.* Characterization of cellular transcriptomic signatures
904 induced by different respiratory viruses in human reconstituted airway epithelia.
905 *Scientific Reports* **9** (2019).

906 17. Steuerman, Y. *et al.* Dissection of influenza infection *in vivo* by single-cell RNA
907 sequencing. *Cell Systems*, 679-691 (2018).

908 18. Terrier, O. *et al.* Cellular transcriptional profiling in human lung epithelial cells
909 infected by different subtypes of influenza A viruses reveals an overall down-
910 regulation of the host p53 pathway. *Virology Journal* **8** (2011).

911 19. Wesolowska-Andersen, A. *et al.* Dual RNA-seq reveals viral infections in
912 asthmatic children without respiratory illness which are associated with changes
913 in the airway transcriptome. *Genome Biology* **18**, 1:17 (2017).

914 20. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature*
915 **550**, 204-217 (2017).

916 21. Everman, J.L. *et al.* Functional genomics of CDHR3 confirms its role in HRV-C
917 infection and childhood asthma exacerbations. *Journal of Allergy and Clinical
918 Immunology* **144**, 962-971 (2019).

919 22. Caliskan, M. *et al.* Rhinovirus wheezing illness and genetic risk of childhood-
920 onset asthma. *New England Journal of Medicine* **368**, 1398-1407 (2013).

921 23. Poole, A. *et al.* Dissecting childhood asthma with nasal transcriptomics
922 distinguishes subphenotypes of disease. *J Allergy Clin Immunol* **133**, 670-678
923 e612 (2014).

924 24. Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and
925 immunosuppression. *Lancet* **395**, 1033-1034 (2020).

926 25. Lukassen S, C.R., Elis R SARS-CoV-2 receptor ACE2 and TMPRSS2 are
927 predominantly expressed in a transient secretory cell type in subsegmental
928 bronchial branches. *biorxiv* (2020).

929 26. Kesic, M.J., Meyer, M., Bauer, R. & Jaspers, I. Exposure to ozone modulates
930 human airway protease/antiprotease balance contributing to increased influenza
931 A infection. *PLoS One* **7**, e35108 (2012).

932 27. Peters, M.C. *et al.* A Transcriptomic Method to Determine Airway Immune
933 Dysfunction in T2-High and T2-Low Asthma. *Am J Respir Crit Care Med* **199**,
934 465-477 (2019).

935 28. Advani, S., Sengupta, A., Forman, M., Valsamakis, A. & Milstone, A.M. Detecting
936 respiratory viruses in asymptomatic children. *Pediatr Infect Dis J* **31**, 1221-1226
937 (2012).

938 29. Jartti, T., Jartti, L., Peltola, V., Waris, M. & Ruuskanen, O. Identification of
939 respiratory viruses in asymptomatic subjects: asymptomatic respiratory viral
940 infections. *Pediatr Infect Dis J* **27**, 1103-1107 (2008).

941 30. Singleton, R.J. *et al.* Viral respiratory infections in hospitalized and community
942 control children in Alaska. *J Med Virol* **82**, 1282-1290 (2010).

943 31. Stelzer-Braend, S. *et al.* Absence of back to school peaks in human rhinovirus
944 detections and respiratory symptoms in a cohort of children with asthma. *J Med
945 Virol* **88**, 578-587 (2016).

946 32. Neophytou, A.M. *et al.* Air Pollution and Lung Function in Minority Youth with
947 Asthma in the GALA II (Genes-Environments and Admixture in Latino

948 Americans) and SAGE II (Study of African Americans, Asthma, Genes, and
949 Environments) Studies. *Am J Respir Crit Care Med* **193**, 1271-1280 (2016).

950 33. Nishimura, K.K. *et al.* Early-life air pollution and asthma risk in minority children.
951 The GALA II and SAGE II studies. *Am J Respir Crit Care Med* **188**, 309-318
952 (2013).

953 34. Thakur, N. *et al.* Socioeconomic status and childhood asthma in urban minority
954 youths. The GALA II and SAGE II studies. *Am J Respir Crit Care Med* **188**, 1202-
955 1209 (2013).

956 35. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI
957 TOPMed Program. *bioRxiv* (2019).

958 36. Everman, J.L., Rios, C. & Seibold, M.A. Utilization of Air-Liquid Interface Cultures
959 as an In Vitro Model to Assess Primary Airway Epithelial Cell Responses to the
960 Type 2 Cytokine Interleukin-13. *Methods Mol Biol* **1799**, 419-432 (2018).

961 37. Jiang, H., Lei, R., Ding, S.W. & Zhu, S. Skewer: a fast and accurate adapter
962 trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*
963 **15**, 182 (2014).

964 38. Wu, T.D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and
965 splicing in short reads. *Bioinformatics* **26**, 873-881 (2010).

966 39. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-
967 throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).

968 40. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and
969 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

970 41. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation
971 network analysis. *BMC Bioinformatics* **9**, 559 (2008).

972 42. Consortium, G.T. *et al.* Genetic effects on gene expression across human
973 tissues. *Nature* **550**, 204-213 (2017).

974 43. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic
975 RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).

976 44. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package
977 for differential expression analysis of digital gene expression data. *Bioinformatics*
978 **26**, 139-140 (2010).

979 45. Alexander, D.H. & Lange, K. Enhancements to the ADMIXTURE algorithm for
980 individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).

981 46. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for
982 complex non-genetic factors in gene expression levels greatly increases power in
983 eQTL studies. *PLoS Comput Biol* **6**, e1000770 (2010).

984 47. Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T. & Delaneau, O. Fast and
985 efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**,
986 1479-1485 (2016).

987 48. Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and
988 analysis. *Nat Commun* **8**, 15452 (2017).

989 49. Mohammadi, P., Castel, S.E., Brown, A.A. & Lappalainen, T. Quantifying the
990 regulatory effect size of cis-acting genetic variation using allelic fold change.
991 *Genome Res* **27**, 1872-1884 (2017).

992 50. Wang, Q., Jia, P. & Zhao, Z. VERSE: a novel approach to detect virus integration
993 in host genomes through reference genome customization. *Genome Med* **7**, 2
994 (2015).

995 51. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat
996 Methods* **9**, 357-359 (2012).

997 52. Kent, W.J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-664
998 (2002).

999 53. Grabherr, M.G. *et al.* Full-length transcriptome assembly from RNA-Seq data
1000 without a reference genome. *Nat Biotechnol* **29**, 644-652 (2011).

1001 54. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local
1002 alignment search tool. *J Mol Biol* **215**, 403-410 (1990).

1003 55. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-
1004 sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

1005 56. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list
1006 enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

1007 57. Travaglini, K.J. *et al.* A molecular cell atlas of the human lung from single cell
1008 RNA sequencing. *biorxiv* (2020).

1009 58. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray
1010 expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).

1011 59. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-
1012 cell transcriptomic data across different conditions, technologies, and species.
1013 *Nat Biotechnol* **36**, 411-420 (2018).

1014 60. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell
1015 RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**,
1016 296 (2019).

1017 61. Waltman, L. & Van Eck, N.J. A smart local moving algorithm for large-scale
1018 modularity-based community detection. *European Physical Journal B* **86**, 471
1019 (2013).

1020 62. McInnes, L. & Healy, J. Uniform Manifold Approximation and Projection for
1021 Dimension Reduction. *ArXiv* (2018).

1022 63. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low
1023 memory requirements. *Nat Methods* **12**, 357-360 (2015).

1024 64. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-
1025 1902 (2019).

1026

1027

1028

1029

1030

1031

1032

1033 Acknowledgements

1034

1035 This work was supported by NIH grants (MAS) U01 HL138626, R01 HL135156, R01
1036 MD010443, R01 HL128439, P01 HL132821, P01 HL107202, R01 HL117004, and DOD
1037 Grant W81WH-16-2-0018.

1038 The Genes-Environments and Admixture in Latino Americans (GALA II) Study and
1039 E.G.B. were supported by the Sandler Family Foundation, the American Asthma
1040 Foundation, the RWJF Amos Medical Faculty Development Program, the Harry Wm.
1041 and Diana V. Hind Distinguished Professor in Pharmaceutical Sciences II, the National
1042 Heart, Lung, and Blood Institute (NHLBI) [R01HL117004, R01HL128439,
1043 R01HL135156, X01HL134589]; the National Institute of Environmental Health Sciences
1044 [R01ES015794]; the National Institute on Minority Health and Health Disparities
1045 (NIMHD) [P60MD006902, R01MD010443], the National Human Genome Research
1046 Institute [U01HG009080] and the Tobacco-Related Disease Research Program [24RT-
1047 0025, 27IR-0030]. MJW was supported by the NHLBI [K01HL140218]. Burchard NIH
1048 Support: T32 GM007546, U01 HL138626, R01 128439, R01 HL141992, R01
1049 HL141845.

1050 Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine
1051 (TOPMed) program was supported by the National Heart, Lung and Blood Institute
1052 (NHLBI). WGS for "NHLBI TOPMed: Gene-Environment, Admixture and Latino
1053 Asthmatics Study" (phs000920) was performed at the New York Genome Center
1054 (3R01HL117004-02S3) and the University of Washington Northwest Genomics Center
1055 (HHSN268201600032I). Centralized read mapping and genotype calling, along with
1056 variant quality metrics and filtering were provided by the TOPMed Informatics Research
1057 Center (3R01HL-117626-02S1; contract HHSN268201800002I). Phenotype

1058 harmonization, data management, sample-identity QC, and general study coordination
1059 were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1,
1060 U01HL-120393, contract HHSN268201800001I). We gratefully acknowledge the studies
1061 and participants who provided biological samples and data for TOPMed.

1062 WGS of part of GALA II was performed by New York Genome Center under The
1063 Centers for Common Disease Genomics of the Genome Sequencing Program (GSP)
1064 Grant (UM1 HG008901). The GSP Coordinating Center (U24 HG008956) contributed to
1065 cross-program scientific initiatives and provided logistical and general study
1066 coordination. GSP is funded by the National Human Genome Research Institute, the
1067 National Heart, Lung, and Blood Institute, and the National Eye Institute.

1068 The authors wish to acknowledge the following GALA II study collaborators: Shannon
1069 Thyne, UCSF; Harold J. Farber, Texas Children's Hospital; Denise Serebrisky, Jacobi
1070 Medical Center; Rajesh Kumar, Lurie Children's Hospital of Chicago; Emerita Brigino-
1071 Buenaventura, Kaiser Permanente; Michael A. LeNoir, Bay Area Pediatrics; Kelley
1072 Meade, UCSF Benioff Children's Hospital, Oakland; William Rodríguez-Cintrón, VA
1073 Hospital, Puerto Rico; Pedro C. Ávila, Northwestern University; Jose R. Rodríguez-
1074 Santana, Centro de Neumología Pediátrica; Luisa N. Borrell, City University of New
1075 York; Adam Davis, UCSF Benioff Children's Hospital, Oakland; Saunak Sen, University
1076 of Tennessee.

1077 The authors acknowledge the families and patients for their participation and thank the
1078 numerous health care providers and community clinics for their support and
1079 participation in GALA II. In particular, the authors thank the recruiters who obtained the

1080 data: Duanny Alva, MD; Gaby Ayala-Rodríguez; Lisa Caine, RT; Elizabeth Castellanos;
1081 Jaime Colón; Denise DeJesus; Blanca López; Brenda López, MD; Louis Martos; Vivian
1082 Medina; Juana Olivo; Mario Peralta; Esther Pomares, MD; Jihan Quraishi; Johanna
1083 Rodríguez; Shahdad Saeedi; Dean Soto; and Ana Taveras.

1084 The content is solely the responsibility of the authors and does not necessarily
1085 represent the official views of the National Institutes of Health.

1086
1087

1088

1089

1090

1091

1092

1093

1094

1095

1096

1097

1098

1099

1100

1101

1102

1103

1104

1105

1106

1107

1108

1109

1110

1111 *Figure Legends*

1112

1113 **Figure 1. ACE2 and TMPRSS2 are expressed by multiple nasal airway epithelial
1114 cell types**

1115 (a) UMAP visualization of cells derived from a human nasal airway epithelial brushing
1116 depicts multiple epithelial and immune cell types identified through unsupervised
1117 clustering.

1118 (b) Normalized expression of ACE2 in epithelial and immune cell types.

1119 (c) Normalized expression of TMPRSS2 in epithelial and immune cell types.

1120

1121 **Figure 2. TMPRSS2 is a mucus secretory network gene regulated by T2
1122 inflammation**

1123 (a) WGCNA identified networks of co-regulated genes related to mucus secretory
1124 function (black), T2 inflammation-induced mucus secretory function (pink), and

1125 canonical T2 inflammation biomarkers (saddle brown). *TMPRSS2* was within the pink
1126 network. Select pathway and cell type enrichments for network genes are shown.

1127 (b) Scatterplot revealing a strong positive correlation between *TMPRSS2* expression
1128 and summary (eigengene) expression of the T2 inflammatory, mucus secretory network.

1129 (c) Scatterplot revealing a strong positive correlation between *TMPRSS2* expression
1130 and summary (eigengene) expression of the canonical T2 inflammation biomarker
1131 network.

1132 (d) Box plots revealing strong upregulation of *TMPRSS2* expression among T2-high
1133 compared to T2-low subjects.

1134 (e) Scatterplot revealing a strong negative correlation between *ACE2* expression and
1135 summary (eigengene) expression of the T2 inflammation mucus secretory network.

1136 (f) Scatterplot revealing a strong negative correlation between *ACE2* expression and
1137 summary (eigengene) expression of the canonical T2 inflammation biomarker network.

1138 (g) Box plots revealing strong downregulation of *ACE2* expression among T2-high
1139 compared to T2-low subjects.

1140

1141 **Figure 3. *ACE2* and *TMPRSS2* expression are both regulated by IL-13 in the**
1142 **mucociliary airway epithelium**

1143 (a) Experimental schematic detailing the timeline for differentiation of basal airway
1144 epithelial cells into a mucociliary airway epithelium and treatment with chronic (10 days)
1145 or acute (72 hours) IL-13 (10ng/ml).

1146 (b) Box plots of count-normalized expression between paired nasal airway cultures
1147 (control/IL-13) revealing strong downregulation of bulk *ACE2* expression with IL-13
1148 treatment. Differential expression results are from DESeq2.

1149 (c) Box plots of count-normalized expression between paired nasal airway cultures
1150 (control/IL-13) revealing strong upregulation of bulk *TMPRSS2* expression with IL-13
1151 treatment. Differential expression results are from DESeq2.

1152 (d) UMAP visualization of cells derived from control and IL-13 stimulated tracheal airway
1153 ALI cultures depict multiple epithelial cell types identified through unsupervised
1154 clustering.

1155 (e) Violin plots of normalized *ACE2* expression across epithelial cell types from tracheal
1156 airway ALI cultures, stratified by treatment (gray = control, red = IL-13). Differential
1157 expression using a Wilcoxon test was performed between control and IL-13-stimulated
1158 cells with significant differences in expression for a cell type indicated by a * ($p < 0.05$).

1159 (f) Violin plots of normalized *TMPRSS2* expression across epithelial cell types from
1160 tracheal airway ALI cultures, stratified by treatment (gray = control, red = IL-13).
1161 Differential expression using a Wilcoxon test was performed between control and IL-13-
1162 stimulated cells with significant differences in expression for a cell type indicated by a *
1163 ($p < 0.05$).

1164

1165 **Figure 4. *ACE2* is an interferon response network gene regulated by respiratory
1166 virus infections**

1167 (a) Scatter plot revealing a strong positive correlation between *ACE2* expression and
1168 summary (eigengene) expression of the cytotoxic immune response network (purple).

1169 (b) Scatterplot revealing a strong positive correlation between *ACE2* expression and
1170 summary (eigengene) expression of the interferon response network (tan).

1171 (c) WGCNA analysis identified networks of co-regulated genes related to cytotoxic
1172 immune response (purple) and interferon response (tan). *ACE2* was within the purple
1173 network. Select pathway and cell type enrichments for network genes are shown.

1174 (d) Box plots of count-normalized expression from GALA II nasal epithelial samples
1175 reveal strong upregulation of *ACE2* expression among interferon-high compared to
1176 interferon-low subjects. Differential expression results are from DESeq2.

1177 (e) Pie graph depicting the percentage of each type of respiratory virus infection found
1178 among GALA II subjects in whom viral reads were found.

1179 (f) Experimental schematic detailing timeline for differentiation of basal airway epithelial
1180 cells into a mucociliary airway epithelium and experimental infection with HRV-A16.

1181 (g) Box plots of count-normalized expression between paired nasal airway cultures
1182 (control/HRV-A16 infected) revealing strong upregulation of *ACE2* expression with
1183 HRV-A16 infection. Differential expression results are from DESeq2.

1184 (h) Box plots of count-normalized expression between paired nasal airway cultures
1185 (control/HRV-A16-infected) revealing no effect of HRVA-16 on *TMRSS2* expression.
1186 Differential expression results are from DESeq2.

1187

1188 **Figure 5. *ACE2* and *TMRSS2* nasal airway expression are regulated by eQTL
1189 variants**

1190 (a) Locuszoom plot of *ACE2* eQTL signals. The lead eQTL variant (rs18160331) is
1191 highlighted with a purple dot. The strength of Linkage Disequilibrium (LD) between

1192 rs18160331 and other variants is discretely divided into five quantiles and mapped into
1193 five colors (dark blue, sky blue, green, orange, and red) sequentially from low LD to high
1194 LD.

1195 (b) Locuszoom plot of *TMPRSS2* eQTL signals. The three independent eQTL variants
1196 (rs1475908, rs2838057, rs74659079) and their LD with other variants (r^2) are
1197 represented by red, blue, and green color gradient respectively.

1198 (c) Box plots of normalized *ACE2* expression among the three genotypes of the lead
1199 *ACE2* eQTL variant (rs18160331). $\log_2 A_{FC} = \log_2$ of the allelic fold change associated
1200 with the variant.

1201 (d) Box plots of normalized *TMPRSS2* expression among the three genotypes of the
1202 lead *TMPRSS2* eQTL variant (rs1475908). $\log_2 A_{FC} = \log_2$ of the allelic fold change
1203 associated with the variant.

1204 (e) Bar plots depicting allele frequencies of the *ACE2* eQTL variant rs18160331 and
1205 *TMPRSS2* eQTL variants (rs1475908, rs2838057, rs74659079) across world
1206 populations. Allele frequency data were obtained from gnomAD v2.1.1.

1207

1208 **Figure 6. Coronavirus infections elicit an enhanced cytotoxic immune response
1209 from the airway epithelium**

1210 (a) Box plots revealing a strong and equivalent upregulation of summary (eigengene
1211 [E_g]) expression for the interferon response network among HRV and CoV-infected
1212 GALA II subjects, compared to uninfected subjects.

1213 (b) Box plots revealing upregulation in summary (eigengene) expression for the
1214 cytotoxic immune response network among HRV-infected GALA II subjects that is even
1215 stronger for the CoV infected group.

1216 (c) Venn Diagram describing the number of differentially expressed genes in HRV and
1217 CoV infected groups compared to the uninfected group, and the extent of their overlap.
1218 For genes differentially expressed in both groups, select enriched pathways and
1219 underlying genes that are highly differentially expressed are shown.

1220 (d) Top upstream regulators predicted by Ingenuity Pathway Analysis to be regulating
1221 the genes that were upregulated in CoV. Enrichment values for these CoV regulators,
1222 using the HRV upregulated genes are also shown.

1223 (e) Heatmap of the \log_2FC in gene expression for CoV and HRV groups when
1224 compared to the uninfected group. Top significantly upregulated genes are shown,
1225 along with *ACE2*, *IL6*, and genes identified as belonging to cytotoxic pathways, which
1226 were enriched within the virally upregulated CoV group DEGs based on IPA canonical
1227 pathway analysis. Color bars indicate which WGCNA network and or IPA canonical
1228 pathway each gene belongs to.

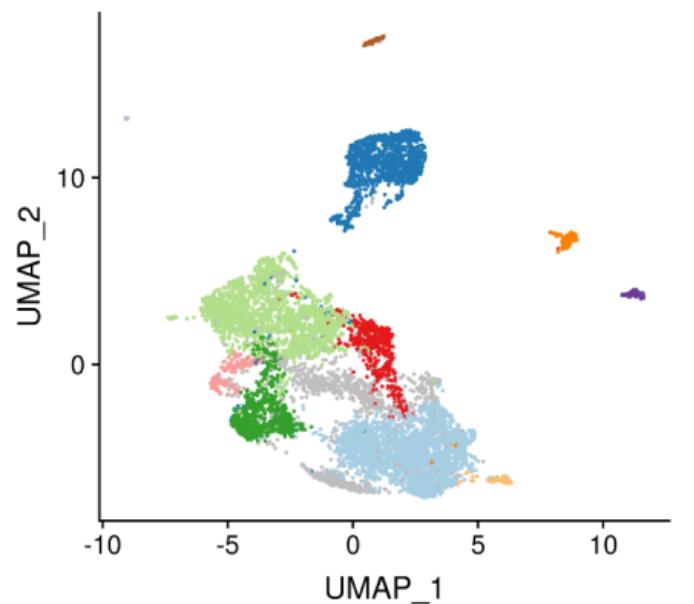
1229 (f) Gene set enrichment analysis plot for CD8+ T cells. The black (shared), yellow (CoV-
1230 enhanced), and red (HRV-enhanced) curves display the enrichment score for the
1231 indicated viral gene set as the analysis walks down the ranked distribution of genes
1232 ordered by fold change in expression between CD8+ T cells relative to all other immune
1233 cell types (red-blue color bar). Genes are represented by vertical bars in the same color
1234 as the curve of the viral gene group they represent. Denoted genes are a representative
1235 set from the leading edge (most responsible for the enrichment).

1236 (g) Gene set enrichment analysis plot for NK cells. The black (shared), yellow (CoV-
1237 enhanced), and red (HRV-enhanced) curves display the enrichment score for the
1238 indicated viral gene set as the analysis walks down the ranked distribution of genes
1239 ordered by fold change in expression between NK cells relative to all other immune cell
1240 types (red-blue color bar). Genes are represented by vertical bars in the same color as
1241 the curve of the viral gene group they represent. Denoted genes are from the leading
1242 edge (most responsible for the enrichment).

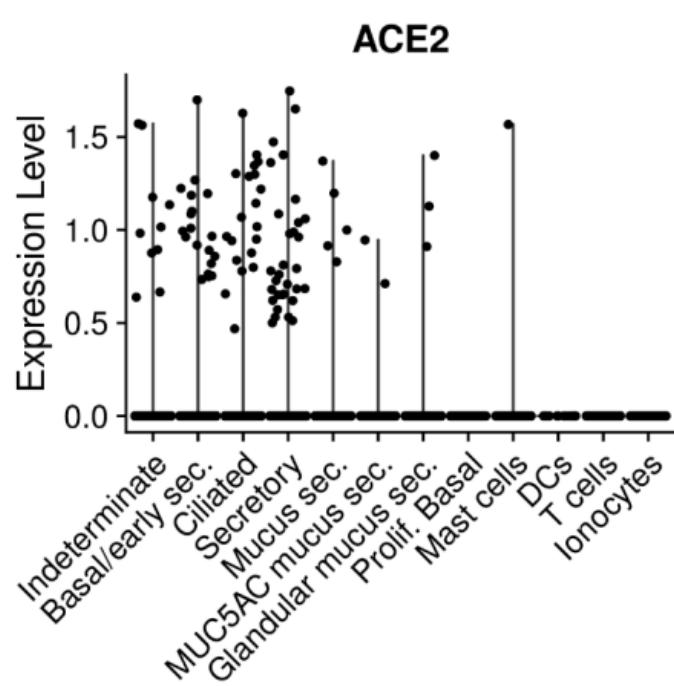
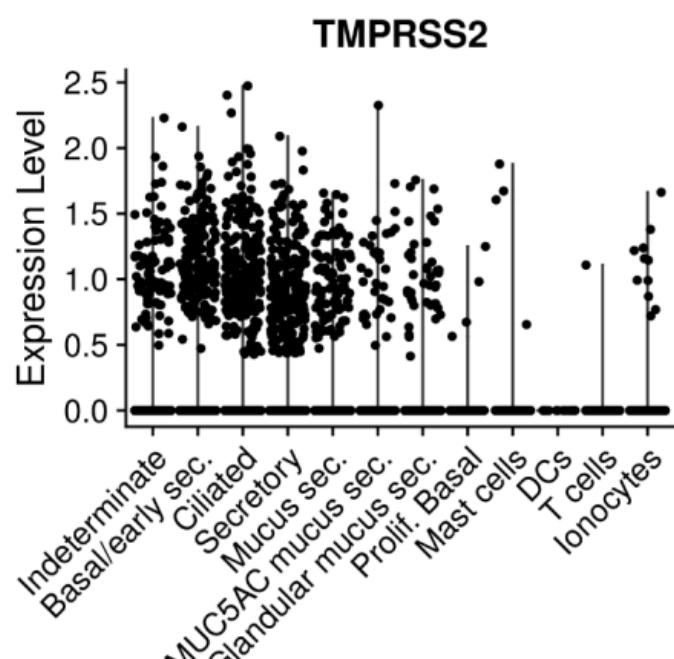
1243

1244

1245

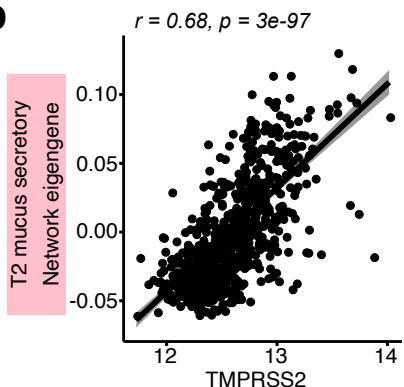
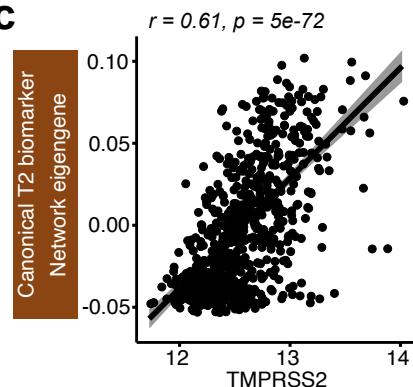
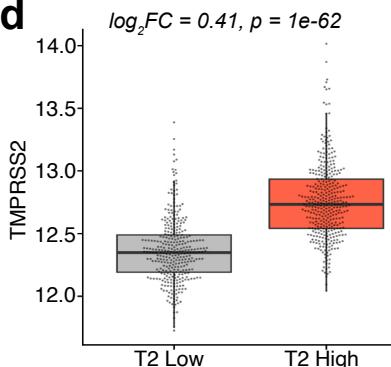
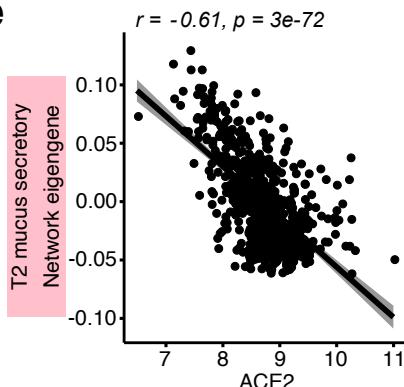
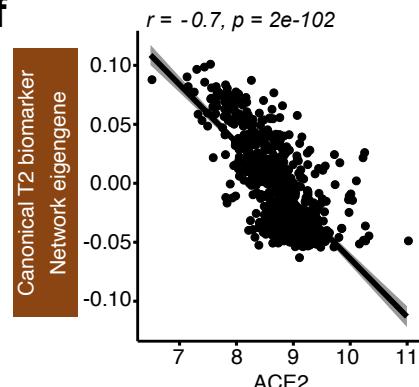
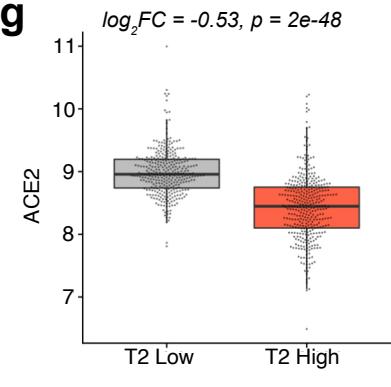
a

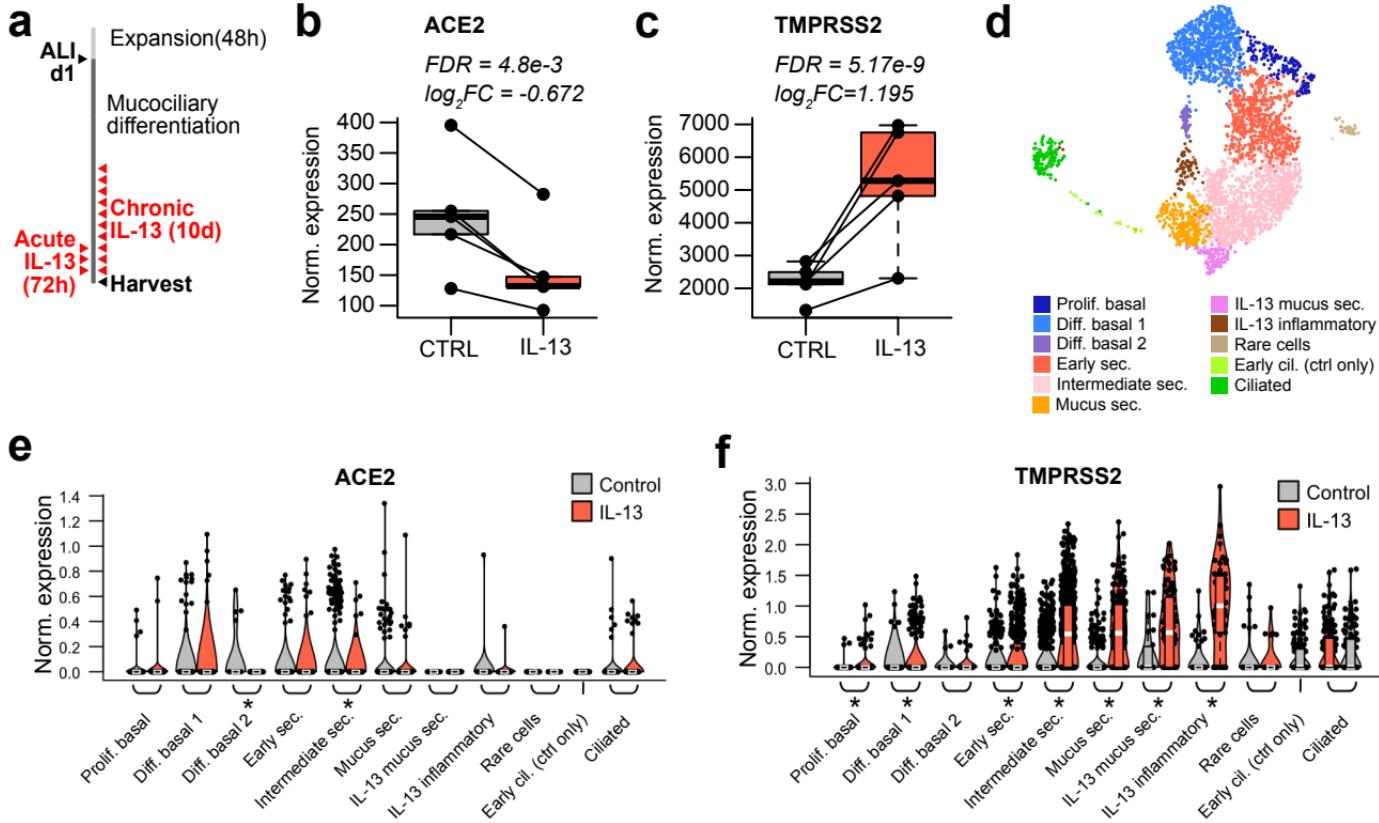
Indeterminate	Glandular mucus sec.
Basal/early sec.	Prolif. Basal
Ciliated	Mast cells
Secretory	DCs
Mucus sec.	T cells
MUC5AC mucus sec.	Ionocytes

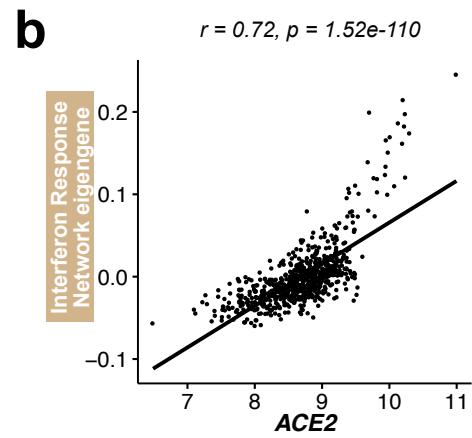
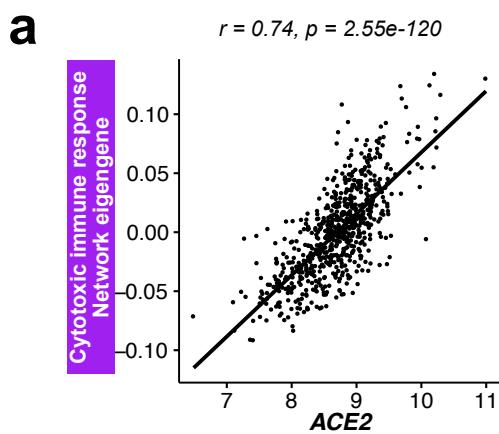
b**c**

a

Network (color)	Network size (# genes)	Select genes	Pathway enrichment	Cell-type enrichment
Mucus secretory (black)	477	COPA COPB2 COPG1 CREB3L1 XBP1	Golgi vesicle transport (<i>p</i> -adj: 2e-6)	Goblet (<i>p</i> -adj: 2e-17)
		SPDEF FCGBP FOXA3 IL13 BPIFB1	COPI-mediated anterograde transport (<i>p</i> -adj: 9e-6)	Diff. basal (<i>p</i> -adj: 0.03)
T2 mucus secretory (pink)	446	CLCA1 CCL26 POSTN IL1RL1 CPA3	O-glycan processing (<i>p</i> -adj: 9e-4)	Goblet (<i>p</i> -adj: 2e-6)
			Polypeptide N-acetylgalactosaminyltransferase activity (<i>p</i> -adj: 3e-3)	Serous (<i>p</i> -adj: 3e-3)
Canonical T2 biomarker (saddle brown)	156	CLCA1 CCL26 POSTN IL1RL1 CPA3	Interleukin-13 human airway epithelial cells (<i>p</i> -adj: 9e-29)	Mast cell/basophil type 1 (<i>p</i> -adj: 6e-11)
			Interleukin-4 human keratinocyte (<i>p</i> -adj: 0.02)	Mast cell/basophil type 2 (<i>p</i> -adj: 6e-11)

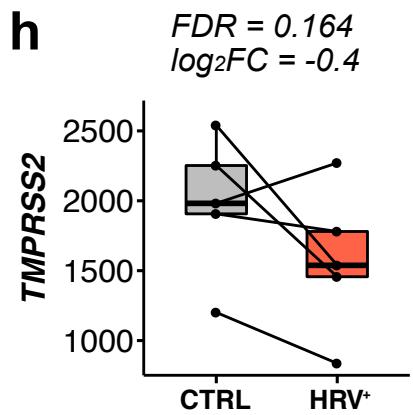
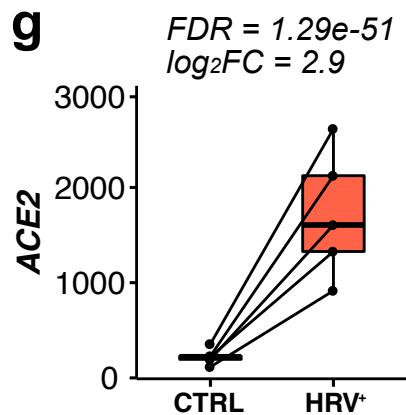
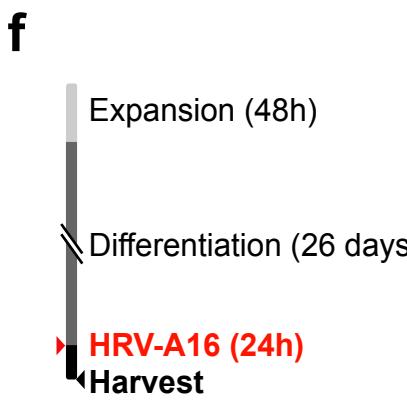
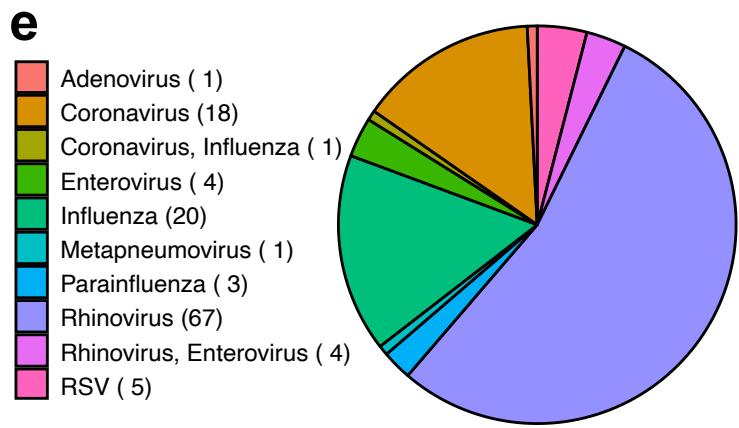
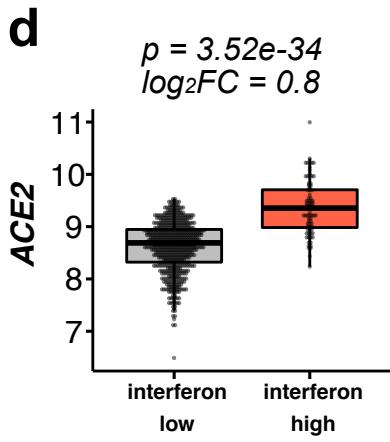
b**c****d****e****f****g**

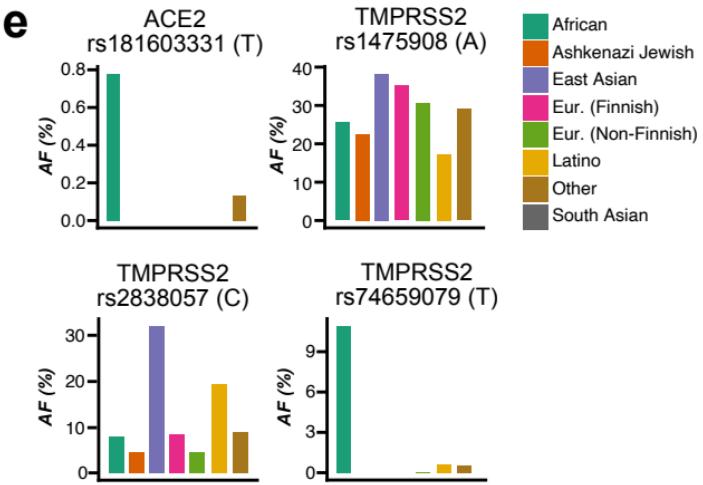
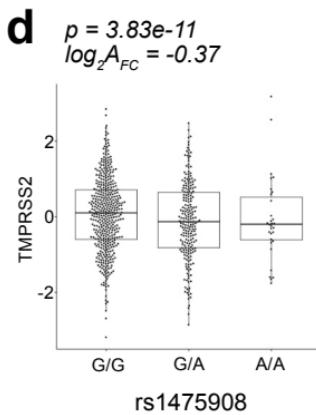
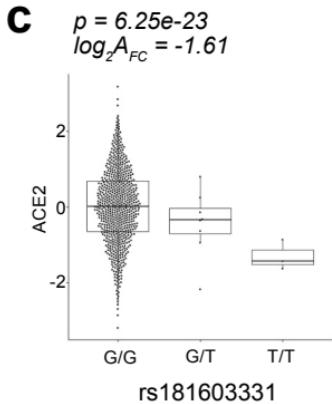
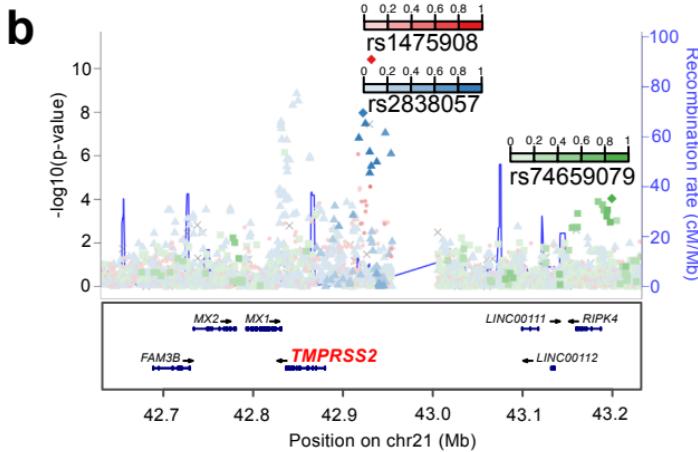
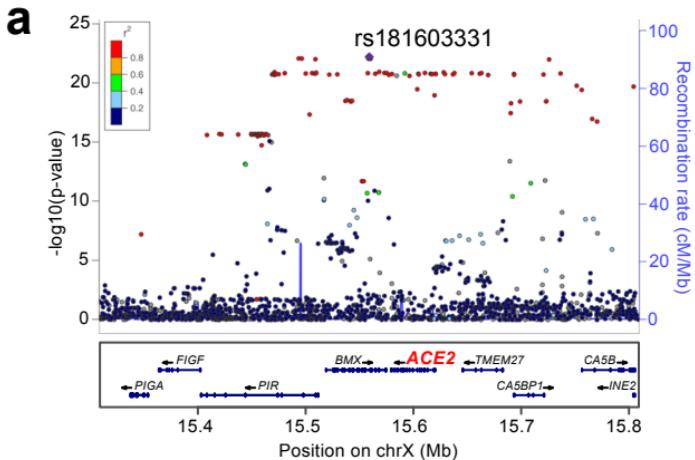


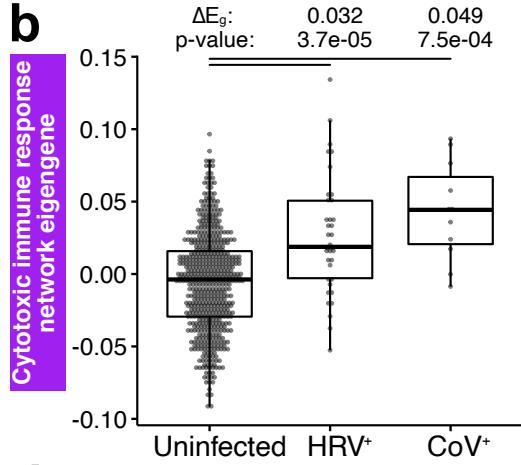
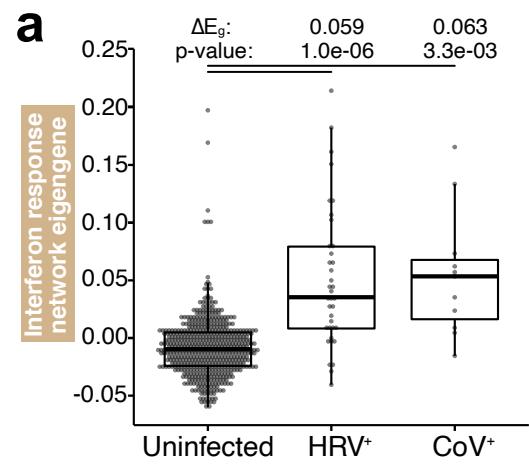


c

Network (color)	Network size (# genes)	Hub genes	Pathway enrichment	Cell-type enrichment
Cytotoxic immune response (purple)	417	IKZF3 CD3E CXCR3 CCR5 NKG7	T cell receptor signaling pathway (<i>p</i> -adj: 6e-16)	CD8 effector T (<i>p</i> -adj: 5e-45)
			Antigen processing and presentation of peptide antigen via MHC class II (<i>p</i> -adj: 1e-7)	Dendritic cell (<i>p</i> -adj: 5e-22)
Interferon response (tan)	296	IFIT2 IFIT3 OAS2 MX1 IRF1	Type I interferon signaling pathway (<i>p</i> -adj: 1e-34)	Monocyte (<i>p</i> -adj: 6e-10)
			AIM2 inflammasome complex (<i>p</i> -adj: 0.05)	Macrophage (<i>p</i> -adj: 1e-8)



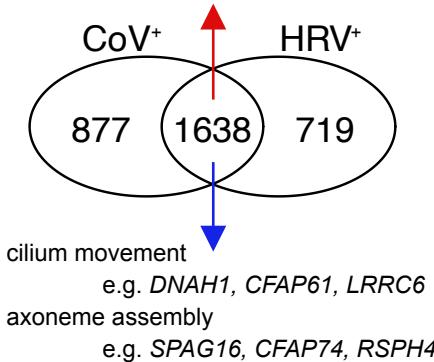




c

cellular response to type I interferon:
e.g. *IFITM1*, *IFI6*, *IFIT1*

inflammatory response:
e.g. *CXCL11*, *IL18RAP*, *TNFRSF8*



d

Upstream Regulator	CoV		HRV	
	Activation Z-score	Enrichment p-value	Activation Z-score	Enrichment p-value
IFNG	11.73	4.04e-110	11.84	9.27e-141
IFNA2	9.04	1.51e-75	8.52	1.40e-74
STAT1	7.88	6.20e-69	7.84	9.33e-83
IFNL1	7.55	2.25e-66	7.28	1.55e-59
TNF	10.16	4.48e-60	11.88	1.74e-98
IRF7	7.89	2.62e-54	8.01	2.40e-57
STAT3	5.00	1.81e-49	6.15	1.72e-77
IL10	2.19	5.44e-49	1.00	1.34e-71
IL1B	8.95	1.64e-47	10.39	4.41e-79
IL6	5.86	1.76e-28	7.32	1.20e-42

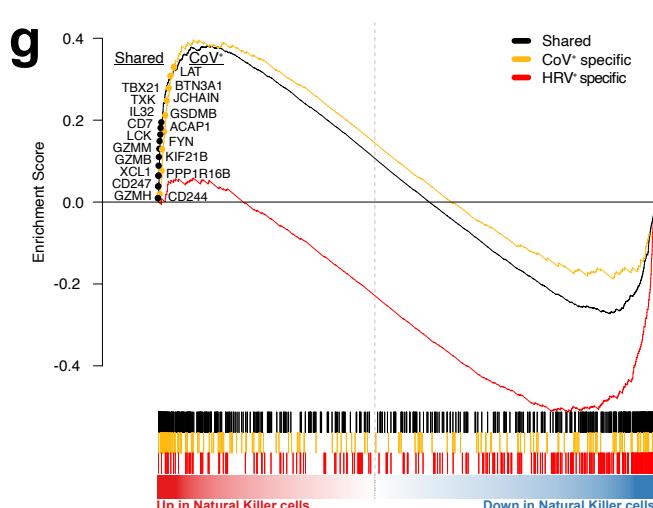
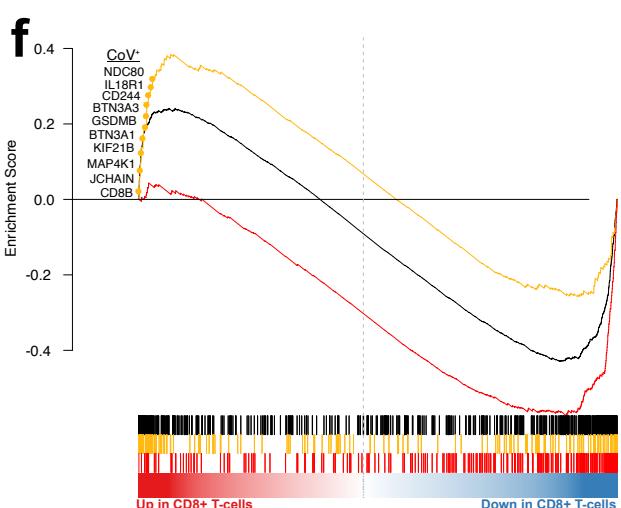
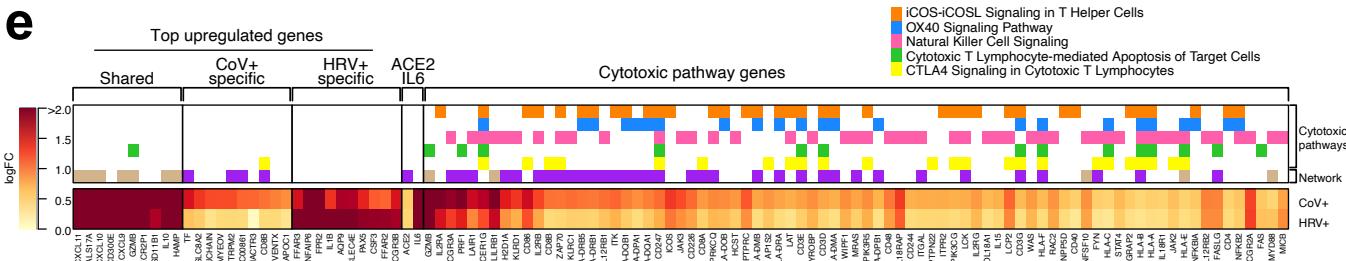


Table 1. Results for multivariate models predicting ACE2 and TMPRSS2 expression

Model	Predictor		Partial R ² (%)	Effect Size		t-test	
	Variable	Ref.		Coeff.	SE	t	p-value
ACE2	Age	n/a	1.03	-0.032	0.009	-3.64	0.000300
	Interferon Status	Low	17.09	1.301	0.088	14.78	6.50e-43
	Type 2 Inflammation	Low	24.44	-1.001	0.057	-17.68	1.58e-57
	Sex	Male	0.14	0.075	0.056	1.33	0.185421
	Asthma	Healthy	0.58	-0.160	0.059	-2.73	0.006415
	rs181603331 (G>T)	G/G	1.20	-0.635	0.162	-3.92	0.000097
TMPRSS2	Age	n/a	0.07	-0.008	0.010	-0.88	0.380112
	Interferon Status	Low	0.07	0.087	0.098	0.88	0.378731
	Type 2 Inflammation	Low	33.24	1.177	0.063	18.77	1.74e-63
	Sex	Male	0.02	-0.031	0.062	-0.50	0.616276
	Asthma	Healthy	<0.01	0.014	0.065	0.22	0.829301
	rs1475908 (G>A)	G/G	0.22	-0.082	0.054	-1.51	0.130251
	rs74659079 (C>T)	C/C	0.39	0.216	0.107	2.03	0.043151
	rs2838057 (A>C)	A/A	0.42	0.139	0.066	2.12	0.034678