



HAL
open science

Environmental and genetic drivers of population differences in SARS-CoV-2 immune responses

Yann Aquino, Aurélie Bisiaux, Zhi Li, Mary O’neill, Javier Mendoza-Revilla, Sarah Hélène Merklings, Gaspard Kerner, Milena Hasan, Valentina Libri, Vincent Bondet, et al.

► **To cite this version:**

Yann Aquino, Aurélie Bisiaux, Zhi Li, Mary O’neill, Javier Mendoza-Revilla, et al.. Environmental and genetic drivers of population differences in SARS-CoV-2 immune responses. 2022. pasteur-03913748

HAL Id: pasteur-03913748

<https://pasteur.hal.science/pasteur-03913748v1>

Preprint submitted on 10 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1 Environmental and genetic drivers of population differences in SARS-CoV- 2 2 immune responses

3

4 Yann Aquino^{1,2,27}, Aurélie Bisiaux^{1,27}, Zhi Li^{1,27}, Mary O'Neill^{1,27}, Javier Mendoza-Revilla¹,
5 Sarah H el ene Merklings³, Gaspard Kerner¹, Milena Hasan⁴, Valentina Libri⁴, Vincent Bondet⁵,
6 Nikaia Smith⁵, Camille de Cevins^{6,7}, Micka el M enager^{6,7}, Francesca Luca^{8,9,10}, Roger Pique-
7 Regi^{8,9}, Giovanna Barba-Spaeth¹¹, Stefano Pietropaoli¹¹, Olivier Schwartz¹², Geert Leroux-
8 Roels¹³, Cheuk-Kwong Lee¹⁴, Kathy Leung^{15,16}, Joseph T.K. Wu^{15,16}, Malik Peiris^{17,18,19},
9 Roberto Bruzzone^{18,19}, Laurent Abel^{20,21,22}, Jean-Laurent Casanova^{20,21,22,23,24}, Sophie A.
10 Valkenburg^{18,25}, Darragh Duffy^{5,19}, Etienne Patin¹, Maxime Rotival^{1,28*} & Llu s Quintana-
11 Murci^{1,26,28*}

12

13 ¹Institut Pasteur, Universit  Paris Cit , CNRS UMR2000, Human Evolutionary Genetics Unit,
14 F-75015 Paris, France.

15 ²Sorbonne Universit , Coll ge Doctoral, Paris, France.

16 ³Institut Pasteur, Universit  Paris Cit , CNRS UMR2000, Insect-Virus Interactions Unit, F-
17 75015 Paris, France.

18 ⁴Institut Pasteur, Universit  Paris Cit , Cytometry and Biomarkers UTechS, F-75015 Paris,
19 France.

20 ⁵Institut Pasteur, Universit  Paris Cit , Translational Immunology U5, F-75015 Paris, France.

21 ⁶Universit  Paris Cit , Imagine Institute, Laboratory of Inflammatory Responses and
22 Transcriptomic Networks in Diseases, Atip-Avenir Team, INSERM UMR1163, 75015 Paris,
23 France

24 ⁷Labtech Single-Cell@Imagine, Imagine Institute, INSERM UMR1163, 75015 Paris, France

25 ⁸Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48201,
26 USA.

27 ⁹Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI 48201,
28 USA.

29 ¹⁰Department of Biology, University of Rome Tor Vergata, 00173 Rome, Italy

30 ¹¹Institut Pasteur, Universit  Paris Cit , CNRS UMR3569, Structural Virology Unit, F-75015
31 Paris, France.

32 ¹²Institut Pasteur, Universit  Paris Cit , CNRS UMR3569, Virus and Immunity Unit, F-75015
33 Paris, France.

34 ¹³Ghent University and University Hospital, Ghent 9000, Belgium.

35 ¹⁴Hong Kong Red Cross Blood Transfusion Service, Hospital Authority, Hong Kong SAR,
36 China.

37 ¹⁵WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of
38 Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong
39 SAR, China.

40 ¹⁶Laboratory of Data Discovery for Health (D²4H), Hong Kong Science Park, Hong Kong
41 SAR, China

42 ¹⁷Division of Public Health Laboratory Sciences, School of Public Health, Li Ka Shing
43 Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.

44 ¹⁸HKU-Pasteur Research Pole, School of Public Health, The University of Hong Kong, Hong
45 Kong SAR, China.

46 ¹⁹Centre for Immunology and Infection, Hong Kong Science Park, Hong Kong SAR, China

47 ²⁰St. Giles Laboratory of Human Genetics of Infectious Diseases, The Rockefeller University,
48 New York, NY, United States.

49 ²¹Laboratory of Human Genetics of Infectious Diseases, INSERM UMR1163, Necker
50 Hospital for Sick Children, Paris, France.

51 ²²Imagine Institute, Paris Cité University, Paris, France.

52 ²³Department of Pediatrics, Necker Hospital for Sick Children, Paris, France.

53 ²⁴Howard Hughes Medical Institute, New York, NY, United States.

54 ²⁵Department of Microbiology and Immunology, Peter Doherty Institute for Infection and
55 Immunity, University of Melbourne, Melbourne, Australia

56 ²⁶Chair Human Genomics and Evolution, Collège de France, Paris 75005, France.

57 ²⁷These authors contributed equally.

58 ²⁸These authors jointly supervised this work.

59 *e-mail: maxime.rotival@pasteur.fr, quintana@pasteur.fr

60

61 **Abstract**

62 Humans display vast clinical variability upon SARS-CoV-2 infection¹⁻³, partly due to genetic
63 and immunological factors⁴. However, the magnitude of population differences in immune
64 responses to SARS-CoV-2 and the mechanisms underlying such variation remain unknown.
65 Here we report single-cell RNA-sequencing data for peripheral blood mononuclear cells from
66 222 healthy donors of various ancestries stimulated with SARS-CoV-2 or influenza A virus.
67 We show that SARS-CoV-2 induces a weaker, but more heterogeneous interferon-stimulated
68 gene activity than influenza A virus, and a unique pro-inflammatory signature in myeloid
69 cells. We observe marked population differences in transcriptional responses to viral exposure
70 that reflect environmentally induced cellular heterogeneity, as illustrated by higher rates of
71 cytomegalovirus infection, affecting lymphoid cells, in African-descent individuals.
72 Expression quantitative trait loci and mediation analyses reveal a broad effect of cell
73 proportions on population differences in immune responses, with genetic variants having a
74 narrower but stronger effect on specific loci. Additionally, natural selection has increased
75 immune response differentiation across populations, particularly for variants associated with
76 SARS-CoV-2 responses in East Asians. We document the cellular and molecular mechanisms
77 through which Neanderthal introgression has altered immune functions, such as its impact on
78 the myeloid response in Europeans. Finally, colocalization analyses reveal an overlap
79 between the genetic architecture of immune responses to SARS-CoV-2 and COVID-19
80 severity. Collectively, these findings suggest that adaptive evolution targeting immunity has
81 also contributed to current disparities in COVID-19 risk.

82 **Introduction**

83 One of the most striking features of the COVID-19 pandemic is the remarkable extent of
84 clinical variation among SARS-CoV-2 infected individuals, ranging from asymptomatic
85 infection to lethal disease¹⁻³. Risk factors include primarily advanced age¹ but also male sex⁵,
86 comorbidities⁶, and human genetic factors (i.e., rare and common variants)^{4,7}. Furthermore,
87 variation in innate immunity⁸⁻¹⁰ – including inborn errors or neutralizing auto-antibodies
88 against type I interferons¹¹⁻¹³ – contribute to the various SARS-CoV-2-related clinical
89 manifestations, and epidemiological and genetic data suggest differences in the outcome of
90 SARS-CoV-2 infection between populations^{6,7,14,15}. These observations, together with
91 previous reports on the importance of ancestry-related differences in transcriptional responses
92 to immune challenges¹⁶⁻¹⁹, highlight the need for in-depth investigations of the magnitude of
93 variation in immune responses to SARS-CoV-2 and its drivers across populations worldwide.

94 There is strong evidence to suggest that viruses and other infectious agents have had an
95 overwhelming impact on human evolution, exerting selection pressures that have shaped
96 present-day population genetic variation²⁰. In particular, human adaptation to RNA viruses,
97 through selective sweeps or admixture with archaic hominins, has been identified as a source
98 of population genetic differentiation^{17,21,22}. For example, strong genetic adaptation, starting
99 ~25,000 years ago, has targeted multiple human coronavirus-interacting proteins in East
100 Asian populations^{23,24}. Furthermore, there is growing evidence for links between Neanderthal
101 introgression and immunity²⁵, with reports of COVID-19 severity being modulated by
102 Neanderthal haplotypes in modern Eurasians^{26,27}. However, the ways in which past natural
103 selection events and archaic admixture have affected the immune response to SARS-CoV-2 in
104 contemporary humans remains to be investigated.

105 We addressed these questions by exposing peripheral blood mononuclear cells (PBMCs)
106 from individuals of Central African, West European, and East Asian descent to SARS-CoV-2
107 and, for the purpose of comparison, to another respiratory RNA virus, influenza A virus
108 (IAV). By combining single-cell RNA sequencing (scRNA-seq) data with quantitative and
109 population genetics approaches, we delineate the respective contributions of cellular, genetic,
110 and evolutionary factors to population variation in immune responses to SARS-CoV-2.

111 **Defining single-cell responses to RNA viruses**

112 We characterized cell type-specific transcriptional responses to SARS-CoV-2 and IAV, by
113 performing scRNA-seq on PBMCs from 222 healthy donors who originate from three
114 different geographic locations (Central Africa, $n = 80$; West Europe, $n = 80$; East Asia, $n =$
115 62 ; Methods), thus exposed to probably different environmental conditions, and carry
116 different genetic ancestries (Supplementary Fig. 1). PBMCs were treated for six hours (i.e., a
117 time point at which immune responses were strong and cell viability was high; Supplementary
118 Fig. 2, Supplementary Note 1, Supplementary Table 1) with a mock-control (non-stimulated),
119 SARS-CoV-2 (ancestral strain, BetaCoV/France/GE1973/2020) or IAV (H1N1/PR/8/1934)
120 ($n=222$ samples for each set of experimental conditions). We captured over one million high-
121 quality single-cell transcriptomes (Fig. 1a, Supplementary Fig. 3, Supplementary Table 2a).
122 By combining transcriptome-based clusters with cellular indexing of transcriptomes and
123 epitopes by sequencing on a subset of samples (CITE-seq; Methods), we defined 22 different
124 cell types across five major immune lineages, including myeloid cells, B cells, CD4⁺ T cells,
125 CD8⁺ T cells and natural killer (NK) cells (Fig. 1b, Supplementary Fig. 4, Supplementary
126 Table 2b-d).

127 After adjusting for technical factors (Methods), we found that cell-type identity was the
128 main driver of gene expression variation (~32%), followed by virus exposure (~27%) (Fig.
129 1b, c, Supplementary Fig. 5). Both SARS-CoV-2 and IAV induced a strong immune response,
130 with 2,914 genes upregulated (FDR < 0.01, $\log_2FC > 0.5$; out of 12,655 with detectable
131 expression) in response to virus stimulation across cell lineages (Supplementary Table 2e).
132 Transcriptional responses to these viruses were highly correlated across cell types and were
133 characterized by a strong induction of interferon-stimulated genes (ISG) (Fig. 1d). However,
134 we observed marked heterogeneity in the myeloid response, with SARS-CoV-2 inducing a
135 specific transcriptional network enriched in inflammatory-response genes (GO:0006954; fold-
136 enrichment (FE) = 3.4, FDR < 4.9×10^{-8}) (Supplementary Table 2f). For example, *IL1A*, *IL1B*
137 and *CXCL8*, encoding pro-inflammatory cytokines, were highly and specifically upregulated
138 in response to SARS-CoV-2 ($\log_2FC > 2.8$, FDR < 2.3×10^{-36}), highlighting the greater
139 inflammatory potential of this virus.

140 We assessed interindividual variability in the response to viral stimuli, by summarizing the
141 response of each individual as a function of their mean level of ISG expression (i.e., ISG
142 activity; Methods, Supplementary Table 2g). We found that SARS-CoV-2 induced more
143 variable ISG activity than IAV across lineages²⁸, with myeloid cells displaying the strongest
144 differences (Supplementary Fig. 6a). We determined the relative contributions of the various
145 interferons (IFNs) to the variation of ISG activity, by using single-molecule arrays (SIMOA)
146 to quantify the levels of secreted IFN- α , β and γ proteins. In the SARS-CoV-2 condition, IFN-
147 α alone accounted for up to 57% of ISG variability, highlighting its determinant role in the
148 response to SARS-CoV-2 (Supplementary Fig. 6b, c). IFN- α transcripts were produced by
149 both infected CD14⁺ monocytes and plasmacytoid dendritic cells (pDCs) after stimulation
150 with IAV, but pDCs were the only source of IFN- α after stimulation with SARS-CoV-2 (Fig.
151 1e) and these cells presented lower levels of *IFNA1-21* expression ($\log_2FC = 6.4$ for SARS-
152 CoV-2 vs. 12.5 for IAV, Wilcoxon rank-sum p -value = 1.2×10^{-16}). Nevertheless, patterns of
153 interindividual variability for ISG activity were remarkably similar after treatment with
154 SARS-CoV-2 and IAV ($r = 0.60$, Pearson's p -value < 1.2×10^{-22} , Fig. 1f), indicating that, at
155 the population level, the IFN-driven response is largely shared between these two viruses.

156 **Marked cellular heterogeneity across populations**

157 We investigated the contribution of differences in cellular proportions to the observed
158 interindividual variability of SARS-CoV-2 responses, by focusing on individuals of Central
159 African and West European ancestries — all recruited during the same sampling campaign,

160 thereby mitigating any potential batch effects related to sample processing¹⁷. We detected
161 marked differences in lineage composition across populations, particularly for NK cells (Fig.
162 2a, Supplementary Table 3a). Notably, an NK subset, identified as memory-like NK cells²⁹,
163 constituted 55.2% of the NK compartment of African-ancestry individuals, but only 12.2% of
164 European-ancestry individuals (Wilcoxon rank-sum p -value $< 6.4 \times 10^{-20}$; Supplementary Fig.
165 7a, b). Individuals of African ancestry also presented higher proportions of CD16⁺
166 monocytes¹⁹ and memory lymphocyte subsets, such as memory B cells, effector CD4⁺ T cells
167 and effector memory CD8⁺ T cells re-expressing CD45RA (CD8⁺ EMRA T cells) (Wilcoxon
168 rank-sum p -value $< 4.1 \times 10^{-5}$).

169 We then searched for genes displaying differences in expression between populations. We
170 found 3,389 genes, across lineages, differentially expressed between populations (popDEGs;
171 FDR < 0.01 , $|\log_2FC| > 0.2$) in the basal state, and 898 and 652 genes displaying differential
172 responses between populations (popDRGs; FDR < 0.01 , $|\log_2FC| > 0.2$) after stimulation with
173 SARS-CoV-2 and IAV, respectively (Fig. 2b, Supplementary Table 3b, c). The popDRGs
174 included genes encoding key immunity regulators, such as the IFN-responsive GBP7 and the
175 macrophage inflammatory protein CCL23 (MIP-3), both of which were more strongly
176 upregulated in Europeans (Fig. 2c). The *GBP7* response was common to both viruses and all
177 lineages ($\log_2FC > 0.88$, Student's t -test adj. p -value $< 1.4 \times 10^{-3}$), but that of *CCL23* was
178 specific to SARS-CoV-2-stimulated myeloid cells ($\log_2FC = 0.72$, Student's adj. p -
179 value $= 5.3 \times 10^{-4}$). We estimated that population differences in cellular composition
180 accounted for 15-47% of popDEGs and for 7-46% of popDRGs, with the strongest impact on
181 NK cells (Fig. 2b, d, Supplementary Fig. 7c). The variation of cellular composition mediated
182 pathway-level differences in response to viral stimulation between populations
183 (Supplementary Table 3d). For example, in virus-stimulated NK cells, genes involved in the
184 promotion of cell migration, such as *CSF1* or *CXCL10*, were more strongly induced in donors
185 of European ancestry (normalized enrichment score > 1.5 , Gene Set Enrichment Analysis adj.
186 p -value < 0.009). However, the loss of this signal after adjustment for cellular composition
187 (Fig. 2e) indicates that fine-scale cellular heterogeneity drives population differences in
188 immune responses to SARS-CoV-2.

189 **Repercussions of latent cytomegalovirus infection**

190 Latent cytomegalovirus (CMV) infection has been reported to vary across populations
191 worldwide³⁰ and to alter cellular proportions³¹⁻³³. We therefore determined the CMV
192 serostatus of the samples. All but one of the individuals of Central African ancestry were

193 CMV⁺ (99%), versus only 31% of donors of West European ancestry, and CMV seropositivity
194 was strongly correlated with the proportions of memory-like NK and CD8⁺ EMRA T cells
195 (Fig. 2f, Supplementary Fig. 7d). Using mediation analysis, we estimated that CMV
196 serostatus accounts for up to 73% of population differences in the proportion of these cell
197 types (Supplementary Table 3e). These differences had a profound impact on the
198 transcriptional response to SARS-CoV-2 (Supplementary Note 2, Supplementary Table 3f),
199 probably contributing to the reported associations between CMV serostatus and COVID-19
200 severity^{34,35}. However, other than its effects on cellular composition, we found that CMV
201 infection had a limited direct effect on the response to SARS-CoV-2, with only one gene
202 presenting significant differences in expression in response to this virus at a FDR of 1%
203 (*ERICH3* in CD8⁺ T cells, log₂FC = 1.7, FDR = 0.007; Supplementary Table 3g). These
204 results reveal how environmental exposures that differ between populations, such as CMV
205 infection, can lead to changes in the composition of the lymphoid fraction that, in turn,
206 explain the observed population differences in the response to SARS-CoV-2.

207 **Genetic architecture of the leukocyte response**

208 We assessed the effects of human genetic variants on transcriptional variation, by mapping
209 expression quantitative trait loci (eQTLs), focusing on *cis*-regulatory variants. At a FDR of
210 1%, we identified between 1,866 and 4,323 independent eQTLs per major cell lineage,
211 affecting a total of 5,198 genes (Fig. 3a, Supplementary Table 4a). Increasing the resolution to
212 22 cell types led to the identification of an additional 3,603 eQTLs (Fig. 3b, Supplementary
213 Fig. 8a, Supplementary Table 4b), highlighting the value of scRNA-seq for identifying
214 context-dependent eQTLs. We found that 79% of eQTLs were replicated (p -value < 0.01) in
215 at least three cell types, but only 22% were common to all lineages. In total, 812 eQTLs were
216 cell type-specific (Methods), ~45% of which were detected in myeloid cells (Fig. 3b),
217 including a pDC-specific eQTL (rs114273142) affecting the host gene encoding miRNA-155
218 — a miRNA that ultimately promotes sensitivity to type I IFNs³⁶ (Supplementary Fig. 8b).
219 More broadly, we found that eQTL effect sizes were highly correlated across ontogenetically
220 related cell types (mean correlation within and between lineages of $r = 0.60$ and 0.47 ,
221 respectively, Wilcoxon rank-sum p -value = 6.2×10^{-6} ; Fig. 3c).

222 We then focused on genetic variants that altered the response to viral stimuli (i.e., response
223 eQTLs, reQTLs). We identified 1,505 reQTLs affecting 1,213 genes (Supplementary Table
224 4c, d). The correlation of effect sizes across ontogenetically related cell types was weaker for
225 reQTLs than for eQTLs (0.36 vs. 0.50, respectively, Wilcoxon rank-sum p -value < 5.6×10^{-13} ,

226 Fig. 3c). Furthermore, the proportion of shared reQTLs between the two viruses differed
227 between cell types. In lymphoid cells, 93% of the reQTLs detected after stimulation with
228 SARS-CoV-2 were also detected after stimulation with IAV (p -value < 0.01), with only 7.7%
229 differing in effect size between viruses (interaction p -value < 0.01 ; Fig. 3d, e). Conversely,
230 the genetic determinants of the myeloid response were much more virus-dependent (49% of
231 myeloid reQTLs, interaction p -value < 0.01), with 46 and 185 reQTLs displaying specific,
232 stronger effects following stimulation with SARS-CoV-2 and IAV, respectively. The
233 strongest SARS-CoV-2-specific reQTL (rs534191, Student's p -value = 1.96×10^{-16} for COV
234 and 0.05 for IAV; Fig. 3f) was identified in myeloid cells, at the *MMP1* locus, which encodes
235 a reported biomarker of COVID-19 severity³⁷. These analyses revealed that the genetic bases
236 of leukocyte responses to SARS-CoV-2 are highly cell type-dependent, with the myeloid
237 response being strongly virus-specific.

238 **Ancestry effects on immune response variation**

239 We then evaluated the contribution of genetic ancestry to population differences in immune
240 responses, by focusing on popDEGs and popDRGs. We found that 11-24% of the genes
241 expressed genome-wide had at least one eQTL, but this proportion increased to up to 56% and
242 60% for popDEGs and popDRGs, respectively, not explained by cellular heterogeneity (Fig.
243 3g, Supplementary Fig. 8c). Furthermore, the popDEGs and popDRGs displaying the largest
244 population differences were more likely to be under genetic control and associated with
245 eQTLs/reQTLs with the largest effect sizes (Supplementary Fig. 8d-f). We used mediation
246 analysis to assess, for each gene, cell lineage and virus treatment, the fraction of population
247 differences explained by genetics (i.e., the most significant eQTL) or cellular heterogeneity
248 (Supplementary Table 5). Cellular composition had a broad effect on population differences
249 in gene expression and responses to viral stimuli (explaining 16-62% of population
250 differences per lineage and virus condition, with the strongest effect in NK cells), whereas
251 genetic variants had a weaker overall effect (accounting for 13-35% of population differences;
252 Fig. 3h, Supplementary Fig. 8g). However, genetic variants had strong effects on a subset of
253 genes (141-433 genes per lineage) for which they accounted for 32-58% of population
254 differences in expression. For example, 81-100% of the difference in *GBP7* expression
255 between donors of African and European ancestry were explained by a single variant
256 displaying strong population differentiation (rs1142888, derived allele frequency (DAF)
257 = 0.13 and 0.53 in African- and European-ancestry individuals, respectively, $F_{ST} = 0.26$,
258 $|\beta_{eQTL}| > 1.7$ across lineages upon stimulation). Thus, variation in immune responses across

259 populations is driven largely by cellular heterogeneity, but common *cis*-genetic variants that
260 present marked allele frequency variation contribute to population differences at specific loci.

261 **Natural selection and SARS-CoV-2 responses**

262 We explored the contribution of natural selection to population differentiation of immune
263 responses. We first searched for overlaps between eQTLs or reQTLs and genome-wide
264 signals of local adaptation, measured by the population branch statistic (PBS)³⁸. We identified
265 1,616 eQTLs (1,215 genes) and 180 reQTLs (166 genes) displaying a strong allele frequency
266 differentiation (empirical *p*-value < 0.01) in at least one population (Supplementary Table
267 S6a). They included key players in IFN-mediated antiviral immunity, such as *DHX58* and
268 *TRIM14* in African-ancestry individuals, *ISG20*, *IFIT5*, *BST2* and *IFITM2-3* in European-
269 ancestry individuals, and *IFI44L* and *IFITM2* in East Asian-ancestry individuals. We then
270 used CLUES³⁹ to identify rapid changes in the frequency trajectory of (r)eQTLs over the last
271 2,000 generations (i.e., 56,000 years) in each population (Supplementary Fig. 9a-d,
272 Supplementary Table S6b). We found signals of rapid adaptation (max. $|Z| > 3$, Methods)
273 targeting the same (*IFITM2*, *IFIT5*) or different (*ISG20*, *IFITM3*, *TRIM14*) eQTLs at highly
274 differentiated genes. We determined whether selection had altered gene expression in specific
275 cell types or in response to SARS-CoV-2 or IAV, by testing for an increase in population
276 differentiation (PBS) at specific eQTLs/reQTLs, relative to random SNPs matched for allele
277 frequency, linkage disequilibrium (LD) and distance to the nearest gene. In the basal state,
278 eQTLs were more strongly differentiated in Europeans, with the strongest signal observed for
279 $\gamma\delta$ T cells (Fig. 4a, Supplementary Fig. 9e). We found that 34% of popDEGs — for which
280 genetics was found to mediate > 50% of the differences between donors of African and
281 European ancestries — were associated with signals of rapid adaptation in Europeans (*vs.*
282 21% in Africans, Fisher's exact *p*-value = 7.7×10^{-6}). For example, population differences at
283 *GBP7* have been driven by a rapid frequency increase, over the last 782-1,272 generations, of
284 the rs1142888-G allele in Europeans (max. $|Z| > 4.3$, Supplementary Fig. 9f).

285 Focusing on the response to viral stimuli, we found that SARS-CoV-2 reQTLs were
286 enriched in signals of population differentiation, specifically in East Asians (fold-enrichment
287 (FE) = 1.24, one-sided resampling *p*-value < 2×10^{-4} , Fig. 4a). Furthermore, among SARS-
288 CoV-2-specific reQTLs, 28 reQTLs (5.3%) displayed signals of rapid adaptation (max. $|Z| >$
289 3) in East Asians starting 770-970 generations ago (~25,000 years) – a time frame associated
290 with polygenic adaptation at SARS-CoV-2-interacting proteins²³ (OR relative to other
291 populations = 2.6, Fisher's exact *p*-value = 7.3×10^{-4} , Fig. 4b, c, Supplementary Fig. 9g, h). A

292 noteworthy example is the immune mediator *LILRB1*, in which we detected a SARS-CoV-2-
293 specific reQTL (rs4806787) in pDCs (Fig. 4d). However, the selection events making the
294 largest contribution to the differentiation of SARS-CoV-2 immune responses in East Asia (top
295 5% PBS) began before this time period (> 970 generations ago, $OR = 1.94$, Fisher's exact p -
296 value = 0.019, Fig. 4b). For example, the rs1028396-T allele, associated with a weaker
297 response of *SIRPA* to SARS-CoV-2 in CD14⁺ monocytes (80% in East Asia vs. 16-25%
298 elsewhere) is characterized by a selection signal beginning more than 45,000 years ago (Fig.
299 4b, e). *SIRPα* has been shown to inhibit infection by endocytic viruses, such as SARS-CoV-2
300 (ref.⁴⁰). These results are consistent with a history of recurrent genetic adaptation targeting
301 antiviral immunity over the last 50,000 years and contributing to present-day population
302 differences in SARS-CoV-2 immune responses.

303 **Functional consequences of Neanderthal introgression**

304 We investigated the effect of the introgression of genetic material from archaic humans, such
305 as Neanderthals or Denisovans, on present-day immune responses to viral challenges, by
306 defining a set of 100,345 introgressed 'archaic' alleles (aSNPs) and determining whether
307 eQTLs were over/underrepresented among introgressed variants relative to random, matched
308 SNPs (Methods). We found that archaic haplotypes were 1.3-1.4 times more likely to alter
309 gene expression, in the basal state (one-sided permutation p -value = 0.02) and after
310 stimulation with SARS-CoV-2 or IAV (one-sided permutation p -value = 5×10^{-4} and 6×10^{-3} ,
311 respectively) in Europeans, whereas this trend was not significant in East Asians ($FE = 1.1$,
312 one-sided permutation p -value > 0.09 , for all sets of conditions, Fig. 5a, Supplementary Table
313 S7a-c). Enrichment was strongest in SARS-CoV-2-stimulated CD16⁺ and IAV-infected
314 CD14⁺ monocytes, suggesting that archaic haplotypes altering myeloid responses to viruses
315 have been preferentially retained in the genomes of modern Europeans. Furthermore, archaic
316 haplotypes regulating gene expression are present at higher frequencies than archaic
317 haplotypes without eQTLs in Europeans, after adjustment for the mean of minor allele
318 frequencies worldwide to ensure similar power for the detection of eQTLs ($\Delta f(\text{introgressed}$
319 $\text{allele}) > 4.1\%$, Student's p -value $< 1.5 \times 10^{-8}$; Methods, Fig. 5b, Supplementary Table S7d, e),
320 providing evidence for the adaptive nature of Neanderthal introgression.

321 We characterized the functional consequences of archaic introgression at the level of
322 individual cell types, by focusing on introgressed eQTLs where the archaic allele was found
323 at its highest frequency in Eurasians (i.e., 5% of most frequent SNPs). These eQTLs included
324 known adaptively introgressed variants at *OAS1-3* or *PNMA1* in Europeans and *TLRI*,

325 *FANCA* or *IL10RA* in East Asians^{17,41-44}, for which we delineated the cellular and molecular
326 effects (Supplementary Fig. 10a, b, Supplementary Table S7f). For example, the COVID-19-
327 associated variant rs10774671 at *OAS1* (ref.⁴⁵) exerts its strongest effect in IAV-stimulated $\gamma\delta$
328 T cells and two alleles in the *TLR1-6-10* region (rs189688666-T and rs112318878-T) have
329 opposite effects on *TLR1* expression in IAV-stimulated monocytes and resting CD4⁺ T cells.

330 We also identified previously unreported signals of Neanderthal introgression affecting
331 immunity phenotypes. For example, an introgressed eQTL (rs11119346-T, 43% in East
332 Asians vs. < 3% in Europeans) was found to downregulate *TRAF3IP3* — encoding a negative
333 regulator of the cytosolic RNA-induced IFN response⁴⁶ — specifically in IAV-infected
334 monocytes, thereby favoring IFN release after viral infection (Supplementary Fig. 10c, d).
335 Likewise, a 35.5-kb Neanderthal haplotype with a frequency of 61% in East Asians (vs. 24%
336 in Europeans) contains the rs9520848-C allele, which is associated with higher basal
337 expression for the cytokine gene *TNFSF13B* in MAIT cells (Supplementary Fig. 10e, f). We
338 also identified an introgressed reQTL (rs58964929-A) at *UBE2F* that was present in 38% of
339 Europeans (vs. 22% of East Asians) and decreased *UBE2F* responses to SARS-CoV-2 and
340 IAV in monocytes (Fig. 5c). *UBE2F* is involved in neddylation, a posttranslational
341 modification required for the nuclear translocation of IRF7 by myeloid cells following
342 infection with RNA viruses and, thus, for the induction of type I IFN responses⁴⁷.
343 Collectively, these results document the molecular and cellular mechanisms through which
344 archaic introgression has altered immune functions.

345 **Immunity-related eQTLs and COVID-19 risk**

346 We investigated the contributions of genetic variants altering responses to SARS-CoV-2 *in*
347 *vitro* to COVID-19 risk *in vivo*, by determining whether eQTLs/reQTLs were more strongly
348 associated with COVID-19 hits detected by genome-wide association studies⁷ than random,
349 matched SNPs (Methods). We observed an enrichment in eQTLs at loci associated with both
350 susceptibility (reported cases) and severity (hospitalized or critical cases) (FE = 4.1 and
351 FE > 3.8, respectively, one-sided resampling *p*-value < 10⁻⁴), and a specific enrichment in
352 reQTLs at severity loci (FE > 3.7, one-sided resampling *p*-value < 3 × 10⁻³; Fig. 6a). This
353 trend was observed across most cell lineages (Supplementary Fig. 11a). Colocalization
354 analyses identified 40 genes at which there was a high probability of (r)eQTL colocalization
355 with COVID-19 hits (coloc. PP_{H4} > 0.8; Supplementary Table S8). These included genes
356 encoding direct regulators of innate immunity, such as *IFNAR2* in non-stimulated CD4⁺ and
357 CD8⁺ T cells, *IRF1* in non-stimulated NK and CD8⁺ T cells, *OAS1* in lymphoid cells

358 stimulated with SARS-CoV-2 and IAV, and *OAS3* in SARS-CoV-2-exposed CD16⁺
359 monocytes (Fig. 6b, c, Supplementary Fig. 11b, c). These results are consistent with a
360 contribution of immunity-related (r)eQTLs to COVID-19 risk.

361 Focusing on the evolutionary factors affecting present-day COVID-19 risk, we identified
362 20 eQTLs that (i) colocalized with COVID-19 susceptibility or severity hits ($PP_{H4} > 0.8$) and
363 (ii) presented positive selection signals (top 1% PBS, $n = 13$ eQTLs) or evidence of archaic
364 introgression ($n = 7$ eQTLs) (Fig. 6d). For example, we found two variants in high LD
365 (rs569414 and rs1559828, $r^2 > 0.73$) at the *DR1* locus that displayed extremely high levels of
366 population differentiation that could be attributed to out-of-Africa selection (DAF = 0.13 in
367 Africa vs. > 0.62 in Eurasia, Supplementary Fig. 11d). Interestingly, DR1 suppresses type I
368 IFN responses⁴⁸ and the alleles subject to selection, which today decrease COVID-19
369 severity, reduce *DR1* expression in most immune cells (Fig. 6d). Furthermore, we identified a
370 ~39-kb Neanderthal haplotype spanning the *MUC20* locus in Europeans and East Asians, in
371 which the rs2177336-T allele is associated with both higher levels of *MUC20* expression in
372 SARS-CoV-2-stimulated cells, particularly for CD4⁺ T cells, and a lower susceptibility to
373 COVID-19. Together, these results reveal the contribution of past selection or Neanderthal
374 introgression affecting immune response variation to current disparities in COVID-19 risk.

375 Discussion

376 The degree and sources of the variation of immune responses to SARS-CoV-2 have emerged
377 as major issues since the beginning of the COVID-19 pandemic^{4,6,14,15}. Based on single-cell
378 approaches, this study provides evidence that cellular proportions, the variation of which is
379 largely due to environmental exposures, are major drivers of population differences in SARS-
380 CoV-2 immune responses. The higher proportions of memory cells detected in lymphoid
381 lineages in individuals of African descent and their association with persistent CMV infection
382 suggest that population differences in cellular activation states may be driven primarily by
383 lifelong pathogen exposure. This highlights how socio-environmental factors (here, pathogen
384 exposure) may covary with individual ancestry (i.e., genetic background), which may lead to
385 an overestimation of the effects of ancestry on phenotypic variation (i.e., immune responses to
386 SARS-CoV-2). Still, common genetic variants can also contribute to the observed variability
387 of immune responses to viral challenges, but their effects tend to be limited to a subset of
388 genes displaying strong population differentiation. This is best illustrated by the rs1142888-G
389 variant, which solely accounts for the > 2.8 -fold higher levels of *GBP7* expression in response
390 to viral stimulation in Europeans than in Africans. The higher frequency of this variant in

391 Europe appears to result from a selection event that occurred 21,900-35,600 years ago. GBP7
392 has been shown to facilitate IAV replication by suppressing innate immunity⁴⁹, but it also
393 regulates IFN- γ -induced oxidative host defense and confers resistance to intracellular bacteria,
394 such as *Listeria monocytogenes* and *Mycobacterium tuberculosis*⁵⁰, providing a plausible
395 mechanism for the occurrence of positive selection at this locus.

396 This study also provides evidence to suggest that past natural selection and admixture
397 with Neanderthals contributed to the differentiation of immune responses to SARS-CoV-2.
398 We found traces of a selection event targeting SARS-CoV-2-specific reQTLs ~25,000 years
399 ago in the ancestors of East Asians, coinciding with the proposed timing of an ancient
400 epidemic affecting the evolution of host coronavirus-interacting proteins^{23,24}. However, we
401 found little overlap between the alleles selected during this period in East Asia and the
402 reported genetic variants underlying COVID-19 risk, suggesting that there have been changes
403 in the genetic basis of infectious diseases over time, possibly due to the evolution of viruses
404 themselves. Nevertheless, we identified cases (e.g., *DRI*, *OAS1-3*, *TOMM7*, *MUC20*) in
405 which selection or archaic introgression contributed to changes in both immune responses to
406 SARS-CoV-2 and the outcome of COVID-19. Genomic studies based on ancestry-aware
407 polygenic risk scores derived from cross-population GWAS will be required to establish a
408 causal link between past adaptation and present-day population differences in COVID-19.

409 Finally, dissection of the genetic architecture of immune response variation across a wide
410 range of cell types provides mechanistic insight into the effect of alleles previously associated
411 with COVID-19 risk. We found that several variants of *IRF1*, *IFNAR2*, and *DRI* associated
412 with lower COVID-19 severity increase type I IFN signaling in lymphoid cells by
413 upregulating *IRF1* and *IFNAR2* or downregulating *DRI*, providing evidence for the
414 importance of efficient IFN signaling for a favorable clinical outcome of SARS-CoV-2
415 infection^{4,11-13}. Another relevant example is provided by the *MUC20* locus, at which we
416 identified a Neanderthal-introgressed eQTL that both increased *MUC20* expression in SARS-
417 CoV-2-stimulated CD4⁺ T cells and has been shown to decrease COVID-19 susceptibility.
418 Given the role of mucins in forming a barrier against infection in the nasal epithelium, we
419 suggest that the Neanderthal haplotype confers greater resistance to viral infections via a
420 similar effect in nasal epithelial cells.

421 Overall, these findings highlight the value of using single-cell approaches to capture the
422 full diversity of the human immune response to RNA viruses, and SARS-CoV-2 in particular,
423 and shed light on the environmental, genetic and evolutionary drivers of immune response
424 variation across individuals and populations.

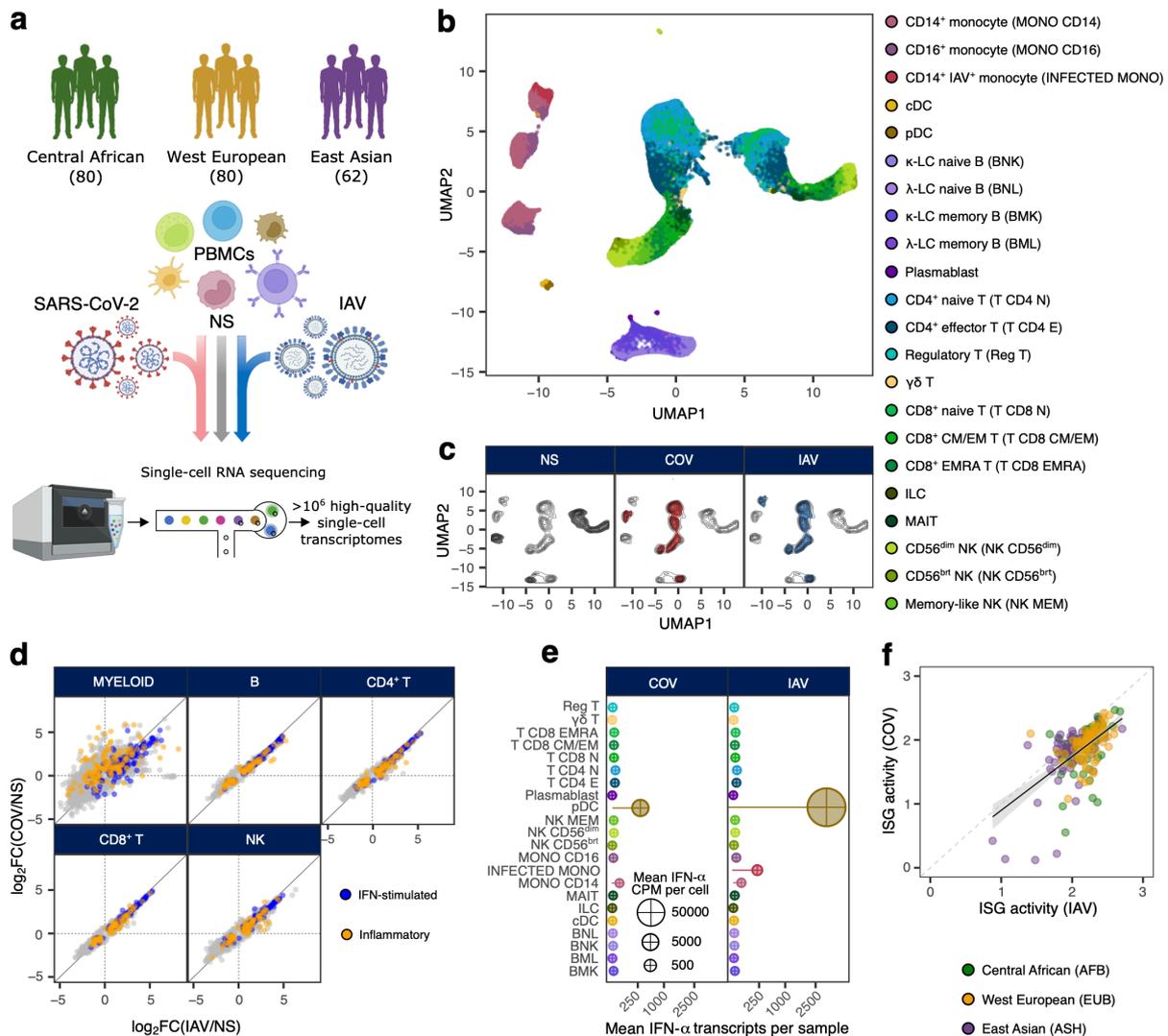
425 **References**

- 426 1. O'Driscoll, M. *et al.* Age-specific mortality and immunity patterns of SARS-CoV-2.
427 *Nature* **590**, 140-145 (2021).
- 428 2. Pei, S., Yamana, T.K., Kandula, S., Galanti, M. & Shaman, J. Burden and
429 characteristics of COVID-19 in the United States during 2020. *Nature* **598**, 338-341
430 (2021).
- 431 3. Sah, P. *et al.* Asymptomatic SARS-CoV-2 infection: A systematic review and meta-
432 analysis. *Proc Natl Acad Sci U S A* **118**, e2109229118 (2021).
- 433 4. Zhang, Q., Bastard, P., Effort, C.H.G., Cobat, A. & Casanova, J.L. Human genetic and
434 immunological determinants of critical COVID-19 pneumonia. *Nature* **603**, 587–598
435 (2022).
- 436 5. Takahashi, T. *et al.* Sex differences in immune responses that underlie COVID-19
437 disease outcomes. *Nature* **588**, 315-320 (2020).
- 438 6. Navaratnam, A.V., Gray, W.K., Day, J., Wendon, J. & Briggs, T.W.R. Patient factors
439 and temporal trends associated with COVID-19 in-hospital mortality in England: an
440 observational study using administrative data. *Lancet Respir Med* **9**, 397-406 (2021).
- 441 7. Kousathanas, A. *et al.* Whole-genome sequencing reveals host factors underlying
442 critical COVID-19. *Nature* **607**, 97-103 (2022).
- 443 8. Schultze, J.L. & Aschenbrenner, A.C. COVID-19 and the human innate immune
444 system. *Cell* **184**, 1671-1692 (2021).
- 445 9. Wilk, A.J. *et al.* A single-cell atlas of the peripheral immune response in patients with
446 severe COVID-19. *Nat Med* **26**, 1070-1076 (2020).
- 447 10. Carvalho, T., Krammer, F. & Iwasaki, A. The first 12 months of COVID-19: a
448 timeline of immunological insights. *Nat Rev Immunol* **21**, 245-256 (2021).
- 449 11. Zhang, Q. *et al.* Inborn errors of type I IFN immunity in patients with life-threatening
450 COVID-19. *Science* **370**, eabd4570 (2020).
- 451 12. Bastard, P. *et al.* Autoantibodies against type I IFNs in patients with life-threatening
452 COVID-19. *Science* **370**, eabd4585 (2020).
- 453 13. Manry, J. *et al.* The risk of COVID-19 death is much greater and age dependent with
454 type I IFN autoantibodies. *Proc Natl Acad Sci U S A* **119**, e2200413119 (2022).
- 455 14. Shelton, J.F. *et al.* Trans-ancestry analysis reveals genetic and nongenetic associations
456 with COVID-19 susceptibility and severity. *Nat Genet* **53**, 801-808 (2021).

- 457 15. Bennett, T.D. *et al.* Clinical Characterization and Prediction of Clinical Severity of
458 SARS-CoV-2 Infection Among US Adults Using Data From the US National COVID
459 Cohort Collaborative. *JAMA Netw Open* **4**, e2116901 (2021).
- 460 16. Nedelec, Y. *et al.* Genetic Ancestry and Natural Selection Drive Population
461 Differences in Immune Responses to Pathogens. *Cell* **167**, 657-669 e21 (2016).
- 462 17. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune
463 System of Human Populations. *Cell* **167**, 643-656 e17 (2016).
- 464 18. Randolph, H.E. *et al.* Genetic ancestry effects on the response to viral infection are
465 pervasive but cell type specific. *Science* **374**, 1127-1133 (2021).
- 466 19. O'Neill, M.B. *et al.* Single-Cell and Bulk RNA-Sequencing Reveal Differences in
467 Monocyte Susceptibility to Influenza A Virus Infection Between Africans and
468 Europeans. *Front Immunol* **12**, 768189 (2021).
- 469 20. Quintana-Murci, L. Human Immunology through the Lens of Evolutionary Genetics.
470 *Cell* **177**, 184-199 (2019).
- 471 21. Enard, D. & Petrov, D.A. Evidence that RNA Viruses Drove Adaptive Introgression
472 between Neanderthals and Modern Humans. *Cell* **175**, 360-371 e13 (2018).
- 473 22. Enard, D. & Petrov, D.A. Ancient RNA virus epidemics through the lens of recent
474 adaptation in human genomes. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190575
475 (2020).
- 476 23. Souilmi, Y. *et al.* An ancient viral epidemic involving host coronavirus interacting
477 genes more than 20,000 years ago in East Asia. *Curr Biol* **31**, 3504-3514 e9 (2021).
- 478 24. Wang, W. & Han, G.Z. Ancient Adaptative Evolution of ACE2 in East Asians.
479 *Genome Biol Evol* **13**, evab173 (2021).
- 480 25. Kerner, G., Patin, E. & Quintana-Murci, L. New insights into human immunity from
481 ancient genomics. *Curr Opin Immunol* **72**, 116-125 (2021).
- 482 26. Zeberg, H. & Paabo, S. The major genetic risk factor for severe COVID-19 is
483 inherited from Neanderthals. *Nature* **587**, 610-612 (2020).
- 484 27. Zeberg, H. & Paabo, S. A genomic region associated with protection against severe
485 COVID-19 is inherited from Neandertals. *Proc Natl Acad Sci U S A* **118**,
486 e2026309118 (2021).
- 487 28. Hadjadj, J. *et al.* Impaired type I interferon activity and inflammatory responses in
488 severe COVID-19 patients. *Science* **369**, 718-724 (2020).
- 489 29. Ram, D.R. *et al.* Tracking KLRC2 (NKG2C)+ memory-like NK cells in SIV+ and
490 rhCMV+ rhesus macaques. *PLoS Pathog* **14**, e1007104 (2018).

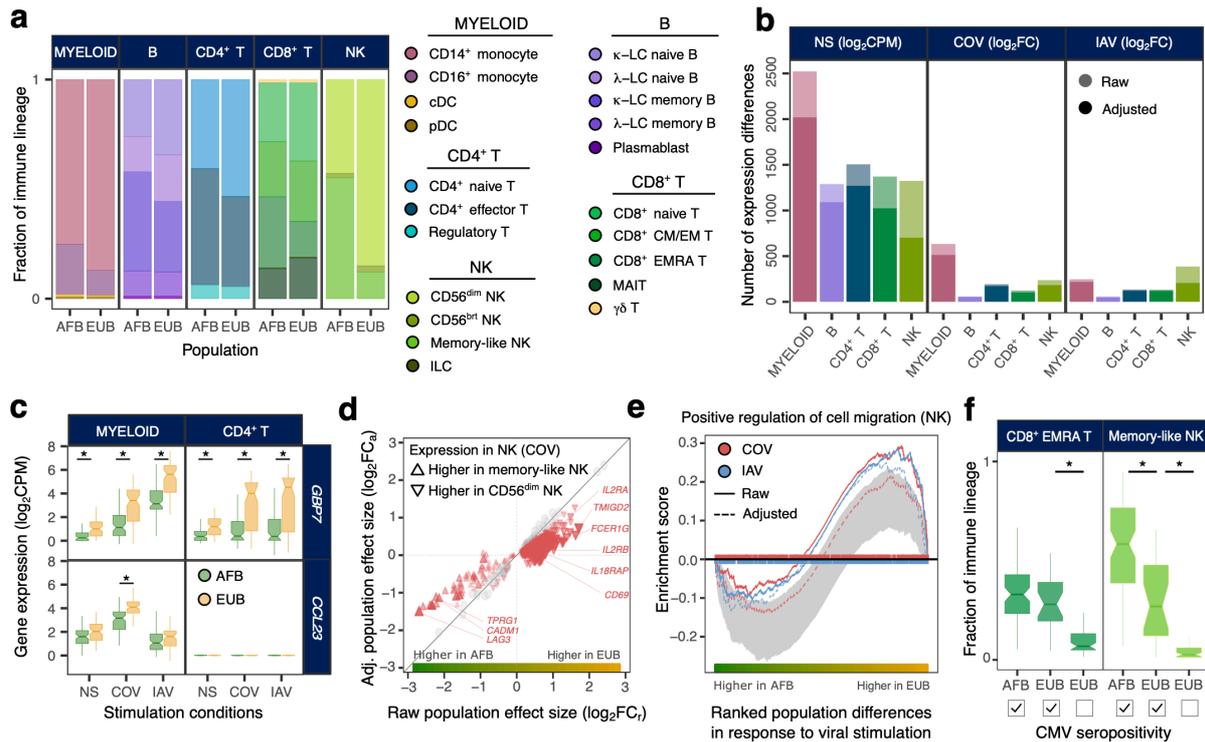
- 491 30. Zuhair, M. *et al.* Estimation of the worldwide seroprevalence of cytomegalovirus: A
492 systematic review and meta-analysis. *Rev Med Virol* **29**, e2034 (2019).
- 493 31. Bigley, A.B., Spielmann, G., Agha, N., O'Connor, D.P. & Simpson, R.J. Dichotomous
494 effects of latent CMV infection on the phenotype and functional properties of CD8+
495 T-cells and NK-cells. *Cell Immunol* **300**, 26-32 (2016).
- 496 32. Guma, M. *et al.* Imprint of human cytomegalovirus infection on the NK cell receptor
497 repertoire. *Blood* **104**, 3664-71 (2004).
- 498 33. Patin, E. *et al.* Natural variation in the parameters of innate immune cells is
499 preferentially driven by genetic factors. *Nat Immunol* **19**, 302-314 (2018).
- 500 34. Alanio, C. *et al.* Cytomegalovirus Latent Infection is Associated with an Increased
501 Risk of COVID-19-Related Hospitalization. *J Infect Dis* **226**, 463-473 (2022).
- 502 35. Weber, S. *et al.* CMV seropositivity is a potential novel risk factor for severe COVID-
503 19 in non-geriatric patients. *PLoS One* **17**, e0268530 (2022).
- 504 36. Wang, P. *et al.* Inducible microRNA-155 feedback promotes type I IFN signaling in
505 antiviral innate immunity by targeting suppressor of cytokine signaling 1. *J Immunol*
506 **185**, 6226-33 (2010).
- 507 37. Syed, F. *et al.* Excessive Matrix Metalloproteinase-1 and Hyperactivation of
508 Endothelial Cells Occurred in COVID-19 Patients and Were Associated With the
509 Severity of COVID-19. *J Infect Dis* **224**, 60-69 (2021).
- 510 38. Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude.
511 *Science* **329**, 75-8 (2010).
- 512 39. Stern, A.J., Wilton, P.R. & Nielsen, R. An approximate full-likelihood method for
513 inferring selection and allele frequency trajectories from DNA sequence data. *PLoS*
514 *Genet* **15**, e1008384 (2019).
- 515 40. Sarute, N. *et al.* Signal-regulatory protein alpha is an anti-viral entry factor targeting
516 viruses using endocytic pathways. *PLoS Pathog* **17**, e1009662 (2021).
- 517 41. Deschamps, M. *et al.* Genomic Signatures of Selective Pressures and Introgression
518 from Archaic Hominins at Human Innate Immunity Genes. *Am J Hum Genet* **98**, 5-21
519 (2016).
- 520 42. Gittelman, R.M. *et al.* Archaic Hominin Admixture Facilitated Adaptation to Out-of-
521 Africa Environments. *Curr Biol* **26**, 3375-3382 (2016).
- 522 43. Racimo, F., Marnetto, D. & Huerta-Sanchez, E. Signatures of Archaic Adaptive
523 Introgression in Present-Day Human Populations. *Mol Biol Evol* **34**, 296-317 (2017).

- 524 44. Choin, J. *et al.* Genomic insights into population history and biological adaptation in
525 Oceania. *Nature* **592**, 583-589 (2021).
- 526 45. Huffman, J.E. *et al.* Multi-ancestry fine mapping implicates OAS1 splicing in risk of
527 severe COVID-19. *Nat Genet* **54**, 125-127 (2022).
- 528 46. Deng, M. *et al.* TRAF3IP3 negatively regulates cytosolic RNA induced anti-viral
529 signaling by promoting TBK1 K48 ubiquitination. *Nat Commun* **11**, 2193 (2020).
- 530 47. Zhao, M. *et al.* Myeloid neddylation targets IRF7 and promotes host innate immunity
531 against RNA viruses. *PLoS Pathog* **17**, e1009901 (2021).
- 532 48. Hsu, S.F., Su, W.C., Jeng, K.S. & Lai, M.M. A host susceptibility gene, DR1,
533 facilitates influenza A virus replication by suppressing host innate immunity and
534 enhancing viral RNA replication. *J Virol* **89**, 3671-82 (2015).
- 535 49. Feng, M. *et al.* Inducible Guanylate-Binding Protein 7 Facilitates Influenza A Virus
536 Replication by Suppressing Innate Immunity via NF-kappaB and JAK-STAT
537 Signaling Pathways. *J Virol* **95**, e02038-20 (2021).
- 538 50. Kim, B.H. *et al.* A family of IFN-gamma-inducible 65-kD GTPases protects against
539 bacterial infection. *Science* **332**, 717-21 (2011).
- 540



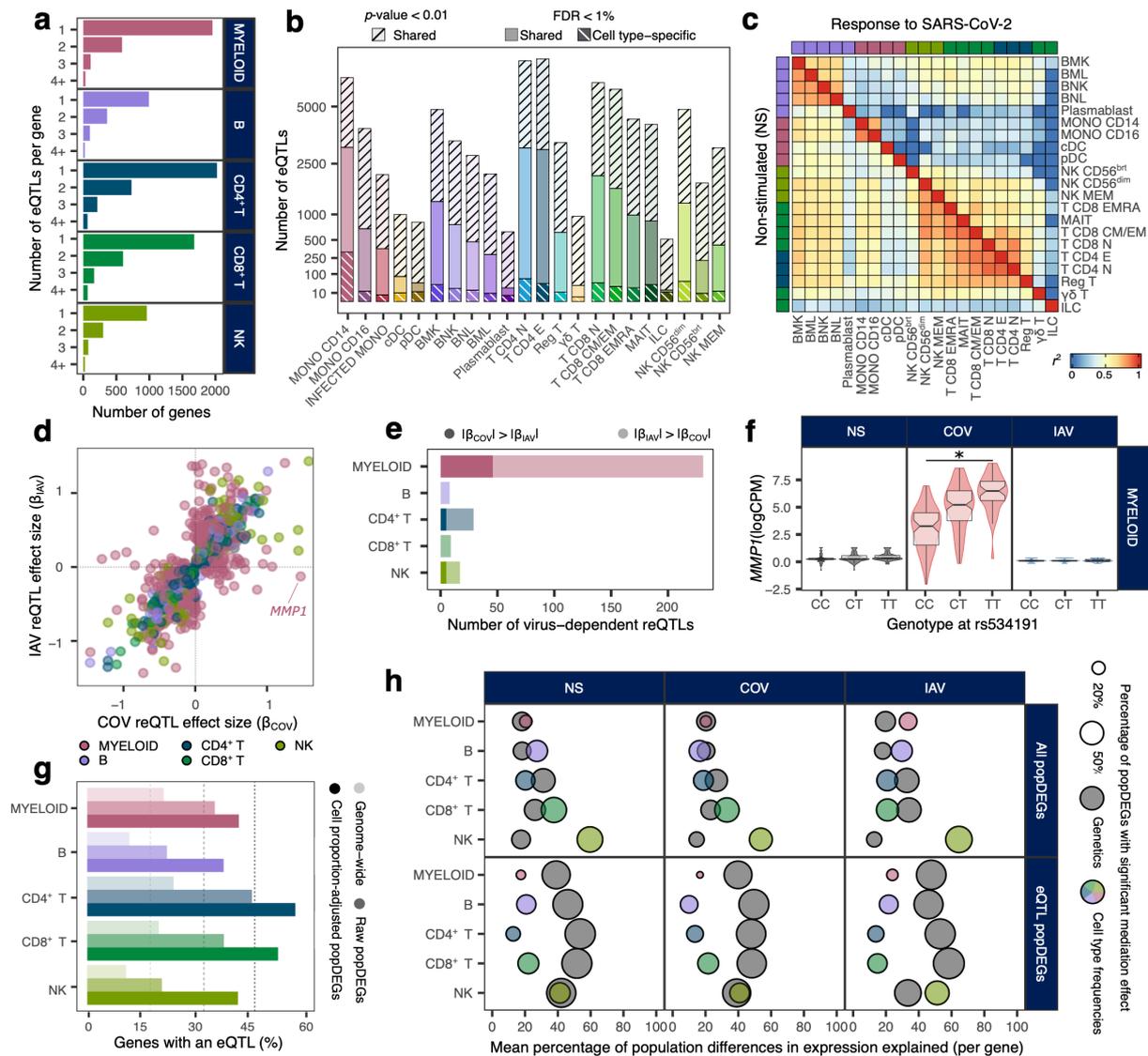
541
542

543 **Figure 1 | Population-scale single-cell responses to SARS-CoV-2 and IAV.** **a**, Study
544 design. **b** and **c**, Uniform manifold approximation and projection (UMAP) of 1,047,824
545 peripheral blood mononuclear cells: resting (non-stimulated; NS), stimulated with SARS-
546 CoV-2 (COV), or influenza A virus (IAV) for six hours. **b**, The colors indicate the 22
547 different cell types inferred. **c**, Distribution of cells in NS, COV and IAV conditions on
548 UMAP coordinates. Contour plot indicates the overall density of cells, and colored areas
549 delineate regions of high cell density in each condition (gray: NS, red: COV, blue: IAV). **d**,
550 Comparison of transcriptional responses to SARS-CoV-2 and IAV across major immune
551 lineages. Hallmark inflammatory and interferon-stimulated genes are highlighted in orange
552 and blue, respectively. **e**, Relative expression of IFN-α-encoding transcripts by each immune
553 cell type in response to SARS-CoV-2 and IAV. Bar lengths indicate the mean number of IFN-
554 α transcripts contributed by each cell type to the overall pool (cell type frequency × mean
555 number of IFN-α transcripts per cell). Dot area is proportional to the mean level of IFN-α
556 transcripts in each cell type (counts per million). **f**, Correlation of ISG activity scores between
557 individuals, following exposure to SARS-CoV-2 and IAV. Each dot corresponds to a single
558 individual (*n* = 222) and its color indicates the self-reported ancestry of the individual
559 concerned (AFB: Central African; EUB: West European; ASH: East Asian).
560



561
562

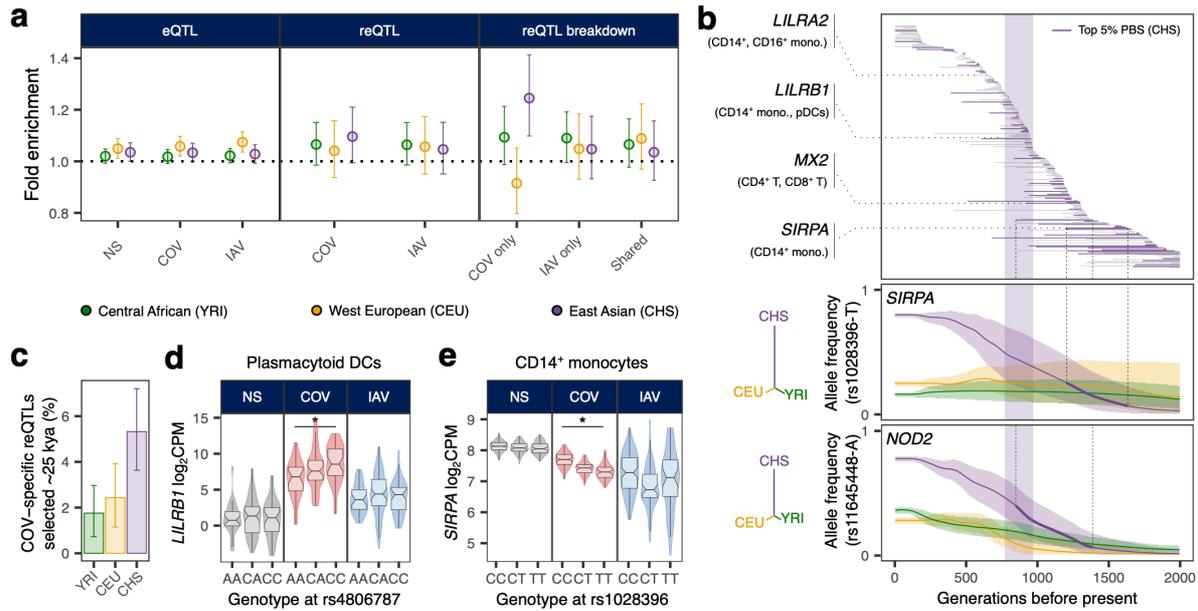
563 **Figure 2 | Effects of the variation of cellular composition on transcriptional responses to**
 564 **viral stimuli.** **a**, Cell-type proportions within each major immune lineage in individuals of
 565 Central African (AFB) and West European (EUB) ancestries. **b**, Number of genes
 566 differentially expressed between AFB and EUB donors, in the basal state (NS) or in response
 567 to SARS-CoV-2 (COV) or influenza A virus (IAV), in each major immune lineage. Numbers
 568 are provided before and after adjustment for cellular composition. **c**, Examples of genes
 569 displaying population differential responses (popDRGs), either shared between cell types and
 570 viruses (*GBP7*) or specific to SARS-CoV-2-stimulated myeloid cells (*CCL23*). *Benjamini-
 571 Hochberg adjusted p -value < 0.001. **d**, Effect of adjusting for cellular composition on genes
 572 differentially expressed between populations. Adjustment reduces the number of genes with
 573 different expression levels between memory-like NK cells and their non-memory counterpart
 574 (i.e., CD56^{dim} NK cells) following exposure to SARS-CoV-2 (red triangles). **e**, Effect of
 575 adjusting for cellular composition on population differences in the response to viral
 576 stimulation for genes involved in the ‘positive regulation of cell migration’ (GO:0030335) in
 577 the NK lineage. **f**, Distribution of CD8⁺ EMRA T and memory-like NK cell frequencies in
 578 AFB and EUB donors according to cytomegalovirus serostatus (CMV^{+/-}) *Wilcoxon Rank-
 579 Sum p -value < 0.001. **d** and **f**, middle line: median; notches: 95% confidence intervals (CI) of
 580 median, box limits: upper and lower quartiles; whiskers: 1.5× interquartile range.
 581



582
583

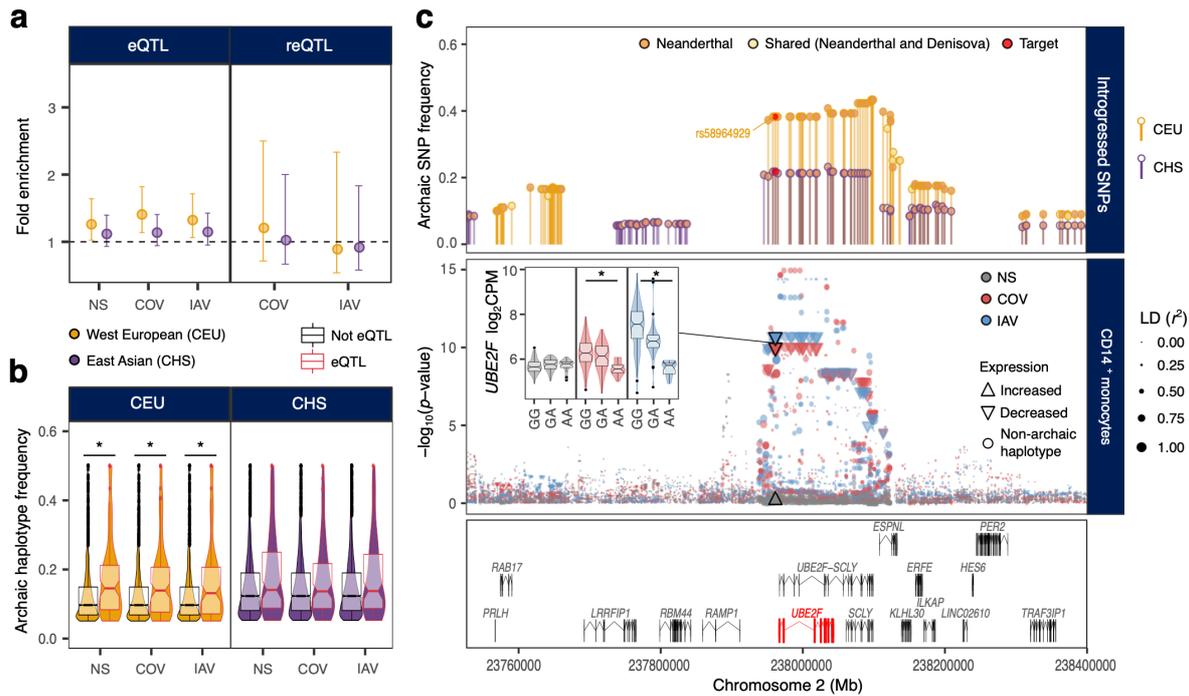
Figure 3 | Genetic architecture of the variation of immune response to RNA viruses. **a**, Number of eQTLs detected per gene within each major immune lineage. **b**, Total number of eQTLs detected in each of the 22 different cell types. Colored bars indicate the number of genome-wide significant eQTLs in each cell type, white stripes (bottom) indicate cell type-specific eQTLs (p -value > 0.01 in all other cell types), and black stripes (top) indicate the total number of eQTLs detected in each cell type including eQTLs from other cell types replicated at a p -value < 0.01 . **c**, Correlation of eQTL (NS; lower triangle) and reQTL (response to SARS-CoV-2; upper triangle) effect sizes across cell types. For each pair of cell types, Spearman's correlation coefficient was calculated for the effect sizes (β) of eQTLs that are significant at a nominal p -value < 0.01 in each cell type. **d**, Comparison of reQTL effect sizes (β) between SARS-CoV-2- and IAV-stimulated cells. Each dot represents a specific reQTL (i.e., SNP, gene, and lineage) and its color indicates the immune lineage in which it was detected. **e**, Number of virus-dependent reQTLs (interaction p -value < 0.01) in each immune lineage, split according to the stimulus where the reQTL has the largest effect size. **f**, Example of a SARS-CoV-2-specific reQTL at the *MMP1* locus. *Student's t -test p -value $< 10^{-16}$; middle line: median; notches: 95% CI of median, box limits: upper and lower quartiles; whiskers: $1.5\times$ interquartile range; points: outliers. **g**, Enrichment in eQTLs among genes differentially expressed between populations (popDEGs). For each immune lineage, bars indicate the percentage of genes with a significant eQTL, at the genome-wide scale and

603 among popDEGs, before or after adjustment for cellular composition. **h**, For each immune
604 lineage and stimulation condition, the x -axis indicates the mean percentage difference in
605 expression between populations mediated by genetics (i.e., the most significant eQTL per
606 gene in each immune lineage and condition) or cellular composition, across all popDEGs
607 (top) or the set of popDEGs associated with a significant eQTL (bottom). The size of the dots
608 reflects the percentage of genes with a significant mediated effect at a FDR of 1%.
609



610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633

Figure 4 | Natural selection effects on population differentiation of immune responses. a, Fold-enrichment in local adaptation signals, defined by the population branch statistic (PBS), and 95% CI, for eQTLs and reQTLs relative to randomly selected variants, matched for minor allele frequency, distance to nearest gene, and LD score. **b.** Estimated periods of selection, over the past 2,000 generations, for 245 SARS-CoV-2 reQTLs with significant rapid adaptation signals in East Asians (CHS) (max. $|Z\text{-score}| > 3$). Variants are ordered in descending order of time to the onset of selection. The area shaded in purple highlights a period corresponding to 770 to 970 generations ago that has been shown to be associated with polygenic adaptation signals at host coronavirus-interacting proteins in East Asians²³. Several immunity-related genes are highlighted (top panel). Allele frequency trajectories of two SARS-CoV-2 reQTLs (rs1028396 at *SIRPA* and rs11645448 at *NOD2*) in Central Africans (YRI, green), West Europeans (CEU, yellow) and East Asians (CHS, purple). Shaded areas indicate the 95% CIs. Dendrograms show the estimated population phylogeny for each eQTL based on PBS (i.e., the branch length between each pair of populations is proportional to $-\log_{10}(1-F_{ST})$). **c.** Percentage of SARS-CoV-2-specific reQTLs presenting selection signals in different populations, between 770 and 970 generations ago, with resampling-based 95% CIs. **d, e,** Examples of SARS-CoV-2-induced reQTLs at *LILRB1* (rs4806787) in plasmacytoid dendritic cells and *SIRPA* (rs1028396) in CD14⁺ monocytes. *Student's *t*-test *p*-value < 0.01; middle line: median; notches: 95% CI of median, box limits: upper and lower quartiles; whiskers: 1.5× interquartile range; points: outliers.



634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655

Figure 5 | Impact of archaic introgression on molecular and cellular phenotypes. a, Enrichment of eQTLs and reQTLs in introgressed haplotypes. For each comparison, the mean observed/expected ratio and the 95% CIs are reported (based on 10,000 resamplings). **b,** For each population and stimulation condition, the frequencies of introgressed haplotypes are compared according to their effects on gene expression (eQTL vs. non-eQTL; *Wilcoxon's p -value < 0.001; middle line: median; notches: 95% CI of median, box limits: upper and lower quartiles; whiskers: 1.5× interquartile range; points: outliers). **c,** Example of adaptive introgression at the *UBE2F* locus (reQTL rs58964929). Upper panel: frequency and nature of archaic alleles across the genomic region of chromosome 2 containing *UBE2F*. Each dot represents an allele of archaic origin, and its color indicates whether it was present in the Vindija Neanderthal genome (orange) or was common to the Vindija Neanderthal and Denisova genomes (light yellow). The eQTL index SNP is shown in red. The frequency in West Europeans (CEU, yellow) and East Asians (CHS, purple) is indicated on the y -axis. Middle panel: monocyte eQTL p -values for SNPs at the *UBE2F* locus, color-coded according to stimulation conditions (gray: non-stimulated (NS), red: SARS-CoV-2-stimulated (COV), blue: IAV-stimulated (IAV)). Each dot represents a SNP and its size (area) is proportional to the LD (r^2) values between the SNP and nearby archaic alleles. For the archaic alleles, arrows indicate the effect of the allele on gene expression. Lower panel: gene structure in the chromosome 2 region, with the *UBE2F* gene highlighted in red.

677 (IAV)). For each eQTL, the target allele is defined as either (i) the derived allele for highly
678 differentiated eQTLs, or (ii) the allele that segregates with the archaic haplotype for
679 introgressed eQTLs in Eurasians. When the ancestral state is unknown, the minor allele is
680 used as a proxy for the derived allele. Note that in some cases (e.g., *OAS1*) the introgressed
681 allele can be present at high frequency in Africa, which is attributed to the reintroduction in
682 Eurasia of an ancient allele by Neanderthals⁴⁵.
683

684 **Methods**

685 **Sample collection**

686 The individuals of self-reported African (AFB) and European (EUB) descent studied are part
687 of the EVOIMMUNOPOP cohort¹⁷. Briefly, 390 healthy male donors were recruited between
688 2012 and 2013 in Ghent (Belgium), thus before the COVID-19 pandemic (188 of self-
689 reported African descent, and 202 of self-reported European descent). Blood was obtained
690 from the healthy volunteers, and the peripheral blood mononuclear cell (PBMC) fraction was
691 isolated and frozen. Inclusion in the current study was restricted to 80 nominally healthy
692 individuals of each ancestry, between 19 and 50 years of age at the time of sample collection.
693 Donors of African descent originated from West Central Africa, with >90% being born in
694 either Cameroon or the Democratic Republic of Congo. For this study, an additional 71
695 individuals of East Asian descent (ASH) were included (62 donors left after quality control,
696 see “Single-cell RNA sequencing library preparation and data processing”). ASH individuals
697 were recruited at the School of Public Health, University of Hong Kong, and were included in
698 a community-based sero-epidemiological COVID-19 study (research protocol number JTW
699 2020.02). Inclusion for the study described here was restricted to nominally healthy ASH
700 individuals (30 men and 41 women) aged between 19 and 65 years of age and seronegative
701 for SARS-CoV-2. Samples were collected at the Red Cross Blood Transfusion Service (Hong
702 Kong) where the PBMC fraction was isolated and frozen.

703 In this study, we refer to individuals of ‘Central African’ (AFB), ‘West European’ (EUB)
704 and ‘East Asian’ (ASH) ancestries to describe individuals who are genetically similar (i.e.,
705 lowest F_{ST} values) to populations from West-Central Africa, Western Europe and East Asia,
706 using the 1,000 Genomes (1KG) Project⁵¹ data as a reference (Supplementary Fig. 1a). Of
707 note, the AFB, EUB and ASH samples present no detectable evidence of recent genetic
708 admixture with populations originating from another continent (e.g., AFB present no traces of
709 recent admixture with EUB).

710 All samples were collected after written informed consent had been obtained from the
711 donors, and the study was approved by the ethics committee of Ghent University (Belgium,
712 no. B670201214647), the Institutional Review Board of the University of Hong-Kong (no.
713 UW 20-132), and the relevant French authorities (CPP, CCITRS and CNIL). This study was
714 also monitored by the Ethics Board of Institut Pasteur (EVOIMMUNOPOP-281297).

715

716

717 **Genome-wide DNA genotyping**

718 The AFB and EUB individuals were previously genotyped at 4,301,332 SNPs, with the
719 Omni5 Quad BeadChip (Illumina, California) with processing as previously described¹⁷. The
720 additional 71 ASH donors were genotyped separately at 4,327,108 SNPs with the Infinium
721 Omni5-4 v1.2 BeadChip (Illumina, California). We updated SNP identifiers based on
722 Illumina annotation files ([https://support.illumina.com/content/dam/illumina-
723 support/documents/downloads/productfiles/humanomni5-4/v1-2/infinium-omni5-4-v1-2-a1-
724 b144-rsids.zip](https://support.illumina.com/content/dam/illumina-support/documents/downloads/productfiles/humanomni5-4/v1-2/infinium-omni5-4-v1-2-a1-b144-rsids.zip)) and called the genotypes of all ASH individuals jointly on GenomeStudio
725 ([https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-
726 design/genomestudio.html](https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html)). We then removed SNPs with (i) no “rs” identifiers or with no
727 assigned chromosome or genomic position ($n = 14,637$); (ii) duplicated identifiers
728 ($n = 5,059$); or (iii) a call rate $< 95\%$ ($n = 10,622$). We then used the 1KG Project Phase 3
729 data⁵¹ as a reference for merging the ASH genotyping data with that of AFB and EUB
730 individuals and detecting SNPs misaligned between the three genotype datasets. Before
731 merging, we removed SNPs that (i) were absent from either the Omni5 or 1KG datasets
732 ($n = 469,535$); (ii) were transversions ($n = 138,410$); (iii) had incompatible alleles between
733 datasets, before and after allele flipping ($n = 1,250$); and (iv) had allele frequency differences
734 of more than 20% between the AFB and Luhya from Webuye, Kenya (LWK) and Yoruba
735 from Ibadan, Nigeria (YRI), or between the EUB and Utah residents with Northern and
736 Western European ancestry (CEU) and British individuals from England and Scotland (GBR),
737 or between the ASH and Southern Han Chinese (CHS) ($n = 777$). Once the data had been
738 merged, we performed principal components analysis (PCA) with *PLINK 1.9* (ref.⁵²) and
739 ensured that the three study populations (i.e., AFB, EUB, and ASH) overlapped with the
740 corresponding 1KG populations, to exclude batch effects between genotyping platforms
741 (Supplementary Fig. 1a). The final genotyping dataset included 3,723,840 SNPs.

742

743 **Haplotype phasing and imputation**

744 After merging genotypes from AFB, EUB and ASH donors, we filtered genotypes for
745 duplicates with *bcftools norm --rm-dup all* (v1.16) (ref.⁵³) and lifted all genotypes over to the
746 human genome assembly GRCh38 with GATK’s (v4.1.2.0) *LiftoverVcf* using the
747 *RECOVER_SWAPPED_ALT_REF=TRUE* option⁵⁴. We then filtered out duplicated variants
748 again before phasing genotypes with *SHAPEIT4* (v4.2.1) (ref.⁵⁵) and imputing missing
749 variants with *Beagle5.1* (version: 18May20.d20) (ref.⁵⁶), treating each chromosome
750 separately. For both phasing and imputation, we used the genotypes of 2504 unrelated

751 individuals from the 1,000 Genomes (1KG) Project Phase 3 data as a reference (downloaded
752 from <ftp://ftp/1000genomes.ebi.ac.uk/vol1/ftp/release20130502> and lifted over to GRCh38)
753 and downloaded genetic maps from the GitHub pages of the associated software (i.e.,
754 *SHAPEIT4* for phasing and *Beagle5.1* for imputation). A third round of duplicate filtering was
755 performed after phasing and before imputation with *Beagle5.1* (version: 18May20.d20)
756 (ref.⁵⁶). Phasing was performed setting *-pbwt-depth=8* and imputation was performed
757 assuming an effective population size (N_e) of 20,000. The quality of imputation was assessed
758 by cross-validation; specifically, we performed 100 independent rounds of imputation
759 excluding 1% of the variants and compared the imputed allelic dosage with the observed
760 genotypes for these variants (Supplementary Fig. 1b, c). The results obtained confirmed that
761 imputation quality was satisfactory, with 98% of common variants (i.e., MAF > 5%) having
762 an $r^2 > 0.8$ for the correlation between observed and imputed genotypes (>95% concordance
763 for 96% of common variants). Following imputation, variants with a MAF < 1% or with a low
764 predicted quality of imputation (i.e., DR2 < 0.9) were excluded, yielding a final dataset of
765 13,691,029 SNPs for downstream analyses.

766

767 **Viruses used in this study**

768 The SARS-CoV-2 reference strain used in this study (BetaCoV/France/GE1973/2020) was
769 supplied by the National Reference Centre for Respiratory Viruses hosted by Institut Pasteur
770 (Paris, France) and headed by Dr. Sylvie van der Werf. The human sample from which the
771 strain was isolated was provided by Dr. Laurent Andreoletti from the Robert Debré Hospital
772 (Paris, France). The influenza A virus strain used in this study (IAV, PR/8, H1N1/1934) was
773 purchased from Charles River laboratories (lot n° #3X051116) and provided in ready-to-use
774 aliquots that were stored at -80°C.

775

776 **SARS-CoV-2 stock production**

777 To produce SARS-CoV-2, we used African green monkey kidney Vero E6 cells that were
778 maintained at 37°C in 5% CO₂ in Dulbecco's minimum essential medium (DMEM) (Sigma-
779 Aldrich) supplemented with 10% fetal bovine serum (FBS, Dutscher) and 1%
780 penicillin/streptomycin (P/S, Gibco, Thermo Fisher Scientific). Vero E6 cells were plated at
781 80% confluence in 150 cm² flasks and infected with SARS-CoV-2 at a multiplicity of
782 infection (MOI) of 0.01 in DMEM supplemented with 2% FBS and 1% P/S. After 1 hour, the
783 inoculum was removed and replaced with DMEM supplemented with 10% FBS, 1% P/S, and
784 cells were incubated for 72 hours at 37°C in 5% CO₂. The cell culture supernatant was

785 collected and centrifuged for 10 min at 3,000 r.p.m to remove cellular debris, and
786 polyethylene glycol (PEG; PEG8000, Sigma-Aldrich) precipitation was performed to
787 concentrate the viral suspension. Briefly, 1 L of viral stock was incubated with 250 mL of
788 40% PEG solution (i.e., 8% PEG final) overnight at 4°C. The suspension was centrifuged at
789 10,000g for 30 minutes at 4°C and the resulting pellet was resuspended in 100 mL of RPMI
790 medium (Gibco, Thermo Fisher Scientific) supplemented with 10% FBS (referred to as R10)
791 and viral aliquots were stored at -80°C. SARS-CoV-2 viral titers were determined by focus-
792 forming unit (FFU) assay as previously described⁵⁷. Briefly, Vero-E6 cells were plated in a
793 96-multiwell plate with 2×10^4 cells per well. The cellular monolayer was infected with serial
794 dilutions (1:10) of viral stock and overlaid with a semi-solid 1.5% carboxymethylcellulose
795 (CMC, Sigma-Aldrich) and 1x MEM medium for 36 hours at 37°C. Cells were then fixed
796 with 4% paraformaldehyde (Sigma-Aldrich), and permeabilized with 1x PBS–0.5% Triton X-
797 100 (Sigma-Aldrich). Infectious foci were stained with a human anti-SARS-CoV-2 Spike
798 antibody (H2-162, Hugo Mouquet’s laboratory, Institut Pasteur) and the corresponding HRP-
799 conjugated secondary antibody (Sigma-Aldrich). Foci were visualized by 3,3’-
800 diaminobenzidine staining solution (DAB, Sigma-Aldrich) staining and counted with the
801 BioSpot suite of a C.T.L. ImmunoSpot S6 Image Analyzer.

802

803 ***In vitro* peripheral blood mononuclear cell stimulation**

804 We performed single-cell RNA-sequencing on SARS-CoV-2-, IAV- and mock-stimulated
805 (referred to as “non-stimulated” condition) PBMCs from healthy donors (80 AFB, 80 EUB
806 and 71 ASH) in 16 experimental runs. We first performed a kinetic experiment (run 1) on
807 samples from 4 AFB and 4 EUB stimulated for 0, 6 and 24 hours to validate our *in vitro*
808 model across different time points (Supplementary Fig. 2, Supplementary Table 1). The 6-
809 hour time point was identified as the optimal time point for the analysis (Supplementary Note
810 1). We then processed the rest of the cohort, over runs 2 through 15. Finally, we reprocessed
811 some samples (run 16) to assess technical variability in our setting and to increase *in silico*
812 cell counts (see ‘Single-cell RNA sequencing library preparation and data processing’
813 section). Ancestry-related batch effects were minimized by scheduling sample processing to
814 ensure a balanced distribution of AFB, EUB and ASH donors within each run.

815 For each run, cryopreserved PBMCs were thawed in a 37°C water bath, transferred to 25
816 mL of R10 medium (i.e., RPMI 1640 supplemented with 10% heat-inactivated fetal bovine
817 serum) at 37°C, and centrifuged at 300g for 10 minutes at room temperature. Cells were
818 counted, re-suspended at 2×10^6 cells/mL in warm R10 in 25cm² flasks, and rested overnight

819 (i.e., 14 hours) at 37°C. The next morning, PBMCs were washed and re-suspended at a
820 density of 3.3×10^6 cells/mL in R10; 120 μ L of a suspension containing 4×10^5 cells from each
821 sample was then plated in a 96-well untreated plate (Greiner Bio-One) for each of the three
822 sets of stimulation conditions. We added 80 μ L of either R10 (non-stimulated), SARS-CoV-2
823 or IAV stock (corresponding to 4×10^5 focus-forming units diluted in R10) to the cells, so as
824 to achieve a multiplicity of infection (MOI) of 1 and an optimal PBMC concentration of
825 2×10^6 cells/mL. Cells were incubated at 37°C for 0, 6 or 24 hours for the kinetic experiment
826 (run 1), and for 6 hours for all subsequent runs (runs 2 to 16), in a biosafety level 3 (BSL-3)
827 facility at Institut Pasteur, Paris. The plates were then centrifuged at 300g for 10 minutes and
828 supernatants were stored at -20°C until use (see ‘Supernatant cytokine assays’ section). All
829 samples from the same run were resuspended in Dulbecco’s phosphate-buffered saline (PBS,
830 Gibco), supplemented with 0.04% bovine serum albumin (BSA, Miltenyi Biotec), and
831 multiplexed in eight pools according to a pre-established study design (Supplementary Fig.
832 3a, Supplementary Table 2a). The cells from each pool were counted with a Cell Countess II
833 automated cell counter (Thermo Fisher Scientific) and cell density was adjusted to 1,000
834 viable cells/ μ L 0.04% BSA in PBS.

835

836 **Single-cell RNA sequencing library preparation and data processing**

837 We generated scRNA-seq cDNA libraries with a Chromium Controller (10X Genomics)
838 according to the manufacturer’s instructions for the Chromium Single Cell 3’ Library and Gel
839 Bead Kits (v3.1). Library quality and concentration were assessed with an Agilent 2100
840 Bioanalyzer and a Qubit fluorometer (Thermo Fisher Scientific). The final products were
841 processed for high-throughput sequencing on a HiSeqX platform (Illumina Inc.).

842 Paired-end sequencing reads from each of the 133 scRNA-seq cDNA libraries (13 libraries
843 from the kinetic experiment and 120 from the population-level study) were independently
844 mapped onto the concatenated human (GRCh38), SARS-CoV-2 (hCoV-
845 19/France/GE1973/2020) and IAV (A/Puerto Rico/8/1934(H1N1)) genome sequences with
846 the *STARsolo* aligner (v2.7.8a) (ref.⁵⁸) (Supplementary Fig. 3b). We obtained a mean of
847 10,785 cell-containing droplets per library, and each droplet was assigned to its sample of
848 origin with *Demuxlet* (v0.1) (ref.⁵⁹), based on the genotyping data available for each
849 individual. Singlet/doublet calls were compared with the output of *Freemuxlet* (v0.1) (ref.⁵⁹)
850 to ensure good agreement (Supplementary Fig. 3c-e). We loaded feature-barcode matrices for
851 all cell-containing droplets identified as singlets by *Demuxlet* in each scRNA-seq library onto
852 a *SingleCellExperiment* (v1.14.1) object⁶⁰. Data from barcodes associated with low-quality or

853 dying cells were removed with a hard threshold-based filtering strategy based on three
854 metrics: cells with fewer than 1,500 total UMI counts, 500 detected features or a
855 mitochondrial gene content exceeding 20% were removed from each sequencing library
856 (Supplementary Fig. 3f). We also discarded samples from nine ASH donors from whom fewer than
857 500 cells were obtained in at least one condition (Supplementary Fig. 3g).

858 We then log-normalized raw UMI counts with a unit pseudocount and library size factors
859 (i.e., number of reads associated with each barcode) were calculated with *quickClusters* and
860 *computeSumFactors* from the *scrn* package (v1.20.1) (ref.⁶⁰). We then calculated the mean
861 and variance of log counts for each gene and broke the variance down into a biological and a
862 technical component with the *fitTrendPoisson* and *modelGeneVarByPoisson* functions of
863 *scrn*. This approach assumes that technical noise is Poisson-distributed and simulates
864 Poisson-distributed data to derive the mean-variance relationship expected in the absence of
865 biological variation. Excess variance relative to the null hypothesis is considered to
866 correspond to the biological variance. We retained only those genes for which the biological
867 variance component was positive with a FDR below 1%. We used this filtered feature set and
868 the technical variance component modeled from the data to run principal components analysis
869 (PCA) with *denoisePCA* from *scrn*, thus discarding later components more likely to capture
870 technical noise. Doublets (i.e., barcodes assigned to cells from different individuals captured
871 in the same droplet) are likely to be in close neighborhoods when projected onto a subspace of
872 the data of lower dimensionality (Supplementary Fig. 3h). We therefore used a *k*-nearest
873 neighbors (*k*-NN) approach to discard cryptic doublets (i.e., barcodes associated to different
874 cells from the same individual captured in the same droplet). Barcodes identified as singlets
875 by *Demuxlet* but having over 5/25 doublet NNs in the PCA space, were re-assigned as
876 doublets and excluded from further analyses.

877 Following data preprocessing, we performed a second round of UMI count normalization,
878 feature selection and dimensionality reduction on the cleaned data, to prevent bias due to the
879 presence of low-quality cells and cryptic doublets. Sequence depth differences were equalized
880 between batches (i.e., sequencing libraries) with *multiBatchNorm* from *batchelor* (v1.8.1) to
881 scale library size factors according to the ratio of mean counts between batches⁶¹
882 (Supplementary Fig. 3i). We accounted for the different mean-variance trends in each batch,
883 by applying *modelGeneVarByPoisson* separately for each sequencing library, and then
884 combining the results for all batches with *combineVar* from *scrn* (ref.⁶⁰). We then bound all
885 133 separate preprocessed feature-barcode matrices into a single merged
886 *SingleCellExperiment* object, log-normalized UMI counts according to the scaled size factors

887 and selected genes with mean log-expression values over 0.01 or a biological variance
888 compartment exceeding 0.001 (Supplementary Fig. 3j). Based on this set of highly variable
889 genes and the variance decomposition, we then performed PCA on the whole data set with
890 *denoisePCA*, and then used *Harmony* (v0.1.0) on the PCs to adjust for library effects⁶².

891

892 **Clustering and cell-type assignment**

893 We performed cluster-based cell-type identification in each stimulation condition, according
894 to a four-step procedure. We first performed low-resolution (*res. parameter*=0.8) shared
895 nearest-neighbors graph-based (k=25) clustering with *FindClusters* from *Seurat* (v4.1.1) with
896 assignment to one of three meta-clusters (i.e., myeloid, B lymphoid and T/NK lymphoid)
897 based on the transcriptional profiles of the cells for canonical markers (e.g., *CD3E-F*, *CD14*,
898 *FCGR3A*, *MS4A1*) (Supplementary Fig. 4a, b). We then performed a second round of
899 clustering at higher resolution (*res. parameter*=3) within each meta-cluster and stimulation
900 condition (Supplementary Fig. 4c). We systematically tested for differential expression
901 between each cluster and the other clusters of the same meta-cluster and stimulation
902 condition. This made it possible to define unbiased markers ($|\log_2FC| \neq 0$, FDR < 0.01) for
903 each cluster (Supplementary Fig. 4d). We then used these expression profiles of these genes
904 to assign each cluster manually to one of 22 different *cell types* (Supplementary Fig. 4e),
905 which, for some analyses, were collapsed into five major immune *lineages*. This step was
906 performed in parallel by three investigators to consolidate consensus assignments. We also
907 used cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) data,
908 generated for a subset of cells (2% of the whole data set), to validate our assignments and
909 redefine clusters presenting ambiguous transcriptional profiles (e.g., memory-like NK cells,
910 Supplementary Fig. 4f).

911 By calling cell types from high-resolution, homogeneous clusters, assigned independently
912 for each lineage and stimulation condition (i.e., non-stimulated, SARS-CoV-2, and IAV), we
913 were able to preserve much of the diversity in our data set, while avoiding potential
914 confounding effects due to the stimulation conditions. However, some clusters were
915 characterized by markers associated with different cell types. Most of these clusters
916 corresponded to mixtures of similar cell types (e.g., the expression of *CD3E*, *CD8A*, *NKG7*,
917 *CD16* suggested a mixture of cytotoxic CD8⁺ T and NK cells) and were consistent with the
918 known cell hierarchy. Other, less frequent clusters expressed a combination of markers
919 usually associated with lineages originating from different progenitors (e.g., *CD3E* and *CD19*,
920 associated with T and B lymphocytes, respectively). These clusters were considered

921 incoherent and discarded. In the fourth and final step of our procedure, we used linear
922 discriminant analysis to resolve the mixtures that were consistent with the established cell
923 hierarchy, to obtain a final cell assignment (Supplementary Fig. 4g, h). For clusters of mixed
924 identity AB, we built a training data set from 10,000 observations sampled from the set of
925 cells called as A or B, preserving the corresponding frequencies of these cells in the whole
926 dataset. We then used a model trained on these data to predict the specific cellular identities
927 within the mixed cluster.

928

929 **Cellular indexing of transcriptomes and epitopes by sequencing**

930 As a means to confirm the identity of specific cell types expressing ambiguous markers at the
931 RNA level, during the last experimental run (run 16), half the cells from each experimental
932 condition were used to perform CITE-seq, according to the manufacturer's instructions (10X
933 Genomics). PBMCs were washed, resuspended in chilled 1% BSA in PBS and incubated with
934 human TruStain FcX blocking solution (BioLegend) for 10 minutes at 4°C. Cells were then
935 stained with a cocktail of TotalSeq™-B antibodies (BioLegend) previously centrifuged at
936 14,000g for 10 minutes (Supplementary Table S2b). The cells were incubated for 30 minutes
937 at 4°C in the dark and were then washed three times. Cell density was then adjusted to 1,000
938 viable cells/μL in 1% BSA in PBS. We generated scRNA-seq libraries and cell protein
939 libraries (L131-L134) with the Chromium Single Cell 3' Reagent Kit (v3.1), using the Feature
940 Barcoding technology for Cell Surface Proteins (10X Genomics).

941

942 **Supernatant cytokine assays**

943 Before protein analysis, sample supernatants were treated in the BSL-3 facility to inactivate
944 the viruses, according to a published protocol for SARS-CoV (ref.⁶³), which we validated for
945 SARS-CoV-2. Briefly, all samples were treated with 1% (v/v) TRITON X100 (Sigma-
946 Aldrich) for 2 hours at room temperature, which effectively inactivated both SARS-CoV-2
947 and IAV. Protein concentration was then determined with a commercial Luminex multi-
948 analyte assay (Biotechne, R&D Systems) and the SIMOA Homebrew assay (Quanterix). For
949 the Luminex assay, we used the XL Performance Kit according to the manufacturer's
950 instructions, and proteins were determined with a Bioplex 200 (Bio-Rad). Furthermore, IFN-
951 α, IFN-γ (duplex) and IFN-β (single-plex) protein concentrations were quantified in SIMOA
952 digital ELISA tests developed as Quanterix Homebrews according to the manufacturer's
953 instructions (<https://portal.quantex.com/>). The SIMOA IFN-α assay was developed with two
954 autoantibodies specific for IFN-α isolated and cloned (Evitria, Switzerland) from two

955 APS1/APECED patients⁶⁴ and covered by patent application WO2013/098419. These
956 antibodies can be used for the quantification of all IFN- α subtypes with a similar sensitivity.
957 The 8H1 antibody clone was used to coat paramagnetic beads at a concentration of 0.3
958 mg/mL for use as a capture antibody. The 12H5 antibody was biotinylated (biotin/antibody
959 ratio = 30:1) and used as the detector antibody, at a concentration of 0.3 μ g/mL. The SBG
960 enzyme for detection was used at a concentration of 150 pM. Recombinant IFN α 17/ α I (PBL
961 Assay Science) was used as calibrator. For the IFN- γ assay, the MD-1 antibody clone
962 (BioLegend) was used to coat paramagnetic beads at a concentration of 0.3 mg/mL for use as
963 a capture antibody. The MAB285 antibody clone (R&D Systems) was biotinylated
964 (biotin/antibody ratio = 40:1) and used as the detector antibody at a concentration of 0.3
965 μ g/mL. The SBG enzyme used for detection was used at a concentration of 150 pM.
966 Recombinant IFN- γ protein (PBL Assay Science) was used as a calibrator. For the IFN- β
967 assay, the 710322-9 IgG1, kappa, mouse monoclonal antibody (PBL Assay Science) was used
968 to coat paramagnetic beads at a concentration of 0.3 mg/mL, for use as a capture antibody.
969 The 710323-9 IgG1 kappa mouse monoclonal antibody was biotinylated (biotin/antibody ratio
970 = 40:1) and used as the detector antibody at a concentration of 1 μ g/mL. The SBG enzyme for
971 detection was used at a concentration of 50 pM. Recombinant IFN- β protein (PBL Assay
972 Science) was used as a calibrator. The limit of detection (LOD) of these assays was 0.8 fg/mL
973 for IFN- α , 20 fg/mL for IFN- γ , and 0.2 pg/mL for IFN- β , considering the dilution factor of
974 10.

975

976 **Flow cytometry**

977 Frozen PBMCs from three AFB (CMV⁺) and six EUB (three CMV⁺, three CMV⁻) donors
978 were thawed and allowed to rest overnight, as previously described. For each donor, 10⁶ cells
979 were resuspended in PBS supplemented with 2% fetal bovine serum and incubated with
980 human Fc blocking solution (BD Biosciences) for 10 minutes at 4°C. Cells were then stained
981 with the following antibodies for 30 minutes at 4°C: CD3 VioGreen (clone BW264/56,
982 Miltenyi Biotec), CD14 V500 (clone M5E2, BD Biosciences), CD57 Pacific Blue (clone
983 HNK-1, Biolegend), NKp46 PE (clone 9E2/NKp46, BD Biosciences), CD16 PerCP-Cy5.5
984 (clone 3G8, BD Biosciences), CD56 APC-Vio770 (clone REA196, Miltenyi Biotec), NKG2A
985 FITC (clone REA110, Miltenyi Biotec), NKG2C APC (clone REA205, Miltenyi Biotec). The
986 cells were then washed and acquired on a MACSQuant cytometer (Miltenyi Biotec), and the
987 data were analyzed with *FlowJo* software (v10.7.1) (ref.⁶⁵).

988

989 **Cytomegalovirus IgG ELISA**

990 We determined the cytomegalovirus (CMV) serostatus of AFB ($n = 80$) and EUB ($n = 80$)
991 donors with a human anti-IgG CMV ELISA kit (Abcam) on plasma samples, according to the
992 manufacturer's instructions.

993

994 **Quantification of batch effects and replicability**

995 Once all the samples had been processed, we used the kBET metric (v0.99.6) (ref.⁶⁶) to assess
996 the intensity of batch effects and to quantify the relative effects of technical and biological
997 variation on cell clustering. This made it possible to confirm that the variation across libraries,
998 and across experimental runs, remained limited relative to the variation across individuals or
999 across conditions (Supplementary Fig. 5a). We used technical replicates to assess the
1000 replicability of our observations across independent stimulations. Agreement was good
1001 between the cell proportions and the interferon-stimulated gene (ISG) activity scores inferred
1002 across independent runs ($r > 0.82$, $p < 7.6 \times 10^{-13}$) (Supplementary Fig. 5b, c).

1003

1004 **Pseudobulk estimation, normalization, and batch correction**

1005 Individual variation in gene expression was quantified at two resolutions: five major immune
1006 *lineages* and 22 *cell types*. We aggregated raw UMI counts from all high-quality single-cell
1007 transcriptomes ($n = 1,047,824$) into bulk expression estimates by summing gene expression
1008 values across all cells assigned to the same lineage/cell type and sample (i.e., individual and
1009 stimulation conditions) using the *aggregateAcrossCells* function of *scuttle* (v1.2.1) (ref.⁶⁷).
1010 We then normalized the raw aggregated UMI counts by library size, generating 3,330 lineage-
1011 wise (222 donors \times 3 sets of conditions \times 5 lineages) and 14,652 cell type-wise (666 samples
1012 \times 22 cell types) pseudobulk counts-per-million (CPM) values, for all genes in our data set.
1013 CPM values were then \log_2 -transformed, with an offset of 1 to prevent non-finite values and
1014 to stabilize variation for weakly expressed genes. Genes with a mean CPM < 1 across all
1015 conditions and lineages/cell types were considered to be non-expressed and were discarded
1016 from further analyses, leading to a final set of 12,667 genes at the lineage level (12,672 genes
1017 when increasing granularity to 22 cell types), including 12 viral transcripts. To quantify the
1018 experimental variation induced by experimental run, library preparation and sequencing, and
1019 remove unwanted batch effects, we first used the *lmer* function of the *lme4* package (v1.1-
1020 27.1) (ref.⁶⁸) to fit a linear model of the following form in each stimulation condition and for
1021 each lineage/cell type:

1022

$$1023 \quad \log(1 + CPM_i) = \alpha + IID_i + LIB_i + RUN_i + FLOW_i + \varepsilon_i \quad (1)$$

1024

1025 where CPM_i is the gene expression in sample i (i.e., one replicate of a given individual and set
1026 of experimental conditions), α is the intercept, $IID_i \sim \mathcal{N}(0, \sigma_{IID}^2)$ captures the effect of the
1027 corresponding individual on gene expression, $LIB_i \sim \mathcal{N}(0, \sigma_{LIB}^2)$ captures the effect of 10X
1028 Genomics library preparation, $RUN_i \sim \mathcal{N}(0, \sigma_{RUN}^2)$ captures the effect of the experimental run,
1029 $FLOW_i \sim \mathcal{N}(0, \sigma_{Flowcell}^2)$ captures the effect of the sequencing flow cell, and ε_i captures
1030 residual variation between samples. We then subtracted the estimated value of the library,
1031 experimental run, and flow cell effects (as provided by the *raneft* function) from the
1032 transformed CPMs of each sample, to obtain batch-corrected CPM values. Finally, we
1033 averaged the batch-corrected CPM values obtained across different replicates for the same
1034 individual and set of stimulation conditions, to obtain final estimates of gene expression.

1035 For each cell type and stimulation condition, an inverse-normal rank-transformation was
1036 applied to the \log_2 CPM of each gene, before testing for differences in gene expression
1037 between populations and mapping expression quantitative trait loci. Within each lineage and
1038 set of stimulation conditions, we ranked, for each gene, the pseudobulk expression values of
1039 all individuals, assigning ranks at random for ties, and replaced each observation with the
1040 corresponding quantile from a normal distribution with the same mean and standard deviation
1041 as the original expression data. This inverse-normal rank-transformation rendered
1042 downstream analyses robust to zero-inflation in the data and outlier values, while maintaining
1043 the rank-transformed values on the same scale as the original data.

1044

1045 **Interferon-stimulated gene activity calculation**

1046 Interferon-stimulated genes (ISGs) strongly respond to both viruses across all lineages/cell
1047 types. We therefore evaluated each donor's ISG expression level at basal state or upon
1048 stimulation with either SARS-CoV-2 or IAV by constructing an "ISG activity" score. For the
1049 human genes in our filtered gene set ($N = 12,655$), we defined as ISGs ($n = 174$) those genes
1050 included in the union of GSEA's hallmark ([https://www.gsea-](https://www.gsea-msigdb.org/gsea/msigdb/genesets.jsp?collection=H)
1051 [msigdb.org/gsea/msigdb/genesets.jsp?collection=H](https://www.gsea-msigdb.org/gsea/msigdb/genesets.jsp?collection=H)) "IFN- α response" and "IFN- γ response"
1052 gene sets, but excluded those from the "inflammatory response" set. We then used
1053 *AddModuleScore* from *Seurat* (v4.1.1) (ref.⁶⁹) to measure ISG activity as the mean
1054 pseudobulk expression level of ISGs in each sample minus the mean expression for a hundred

1055 randomly selected non-ISGs matched for mean magnitude of expression. In all analyses, ISG
1056 activity scores were adjusted for cell mortality of the sample by fitting a model of the form:

1057

$$1058 \quad ISG_i = \alpha + \text{Population}_i + \text{CellMortality}_i + \varepsilon_i \quad (2)$$

1059

1060 and subtracting the effect of cell mortality from the raw ISG scores. In this model, ISG_i
1061 denotes the ISG activity score of individual i , α is the intercept, Population_i and
1062 CellMortality_i are variables capturing the effect of the population, and cell mortality on ISG
1063 activity and ε_i are normally distributed residuals. For comparisons with SIMOA-estimated
1064 IFN levels, the *carscore* function from the *care* R package⁷⁰ was used to model ISG activity
1065 as a function of levels of IFN- α , β , and γ , adjusting for population, age, sex and cell
1066 mortality. The percentage of ISG variance attributable to each IFN (α , β , or γ) was estimated
1067 as the square of the resulting CAR scores.

1068

1069 **Modelling population effects on the variation of gene expression**

1070 To estimate population effects on gene expression while mitigating any potential batch effect
1071 relating to sample processing, we first focused exclusively on AFB and EUB individuals, as
1072 all these individuals were recruited during the same sampling campaign and their PBMCs
1073 were processed at the same time, with the same experimental procedure¹⁷. For each immune
1074 lineage, cell type, stimulation condition, and gene, we then built a separate linear model of the
1075 form:

1076

$$1077 \quad \text{Expr}_i = \alpha + \beta_r \cdot I_i^{EUB} + Z_i^T \cdot \gamma + \varepsilon_i \quad (3)$$

1078

1079 where Expr_i is the rank-transformed gene expression (log-normalized CPM) for individual i
1080 in the lineage/cell type and condition under consideration, I_i^{EUB} is an indicator variable equal
1081 to 1 for European-ancestry individuals and 0 otherwise, and Z_i represents the set of core
1082 covariates of the sample that includes the individual's age and cellular mortality (i.e.,
1083 proportion of dying cells in each thawed vial, as a proxy of sample quality). In addition, ε_i are
1084 the normally distributed residuals and α, β_r, γ are the fitted parameters of the models. In
1085 particular, α is the intercept, β_r indicates the \log_2 fold change difference in expression
1086 between individuals of European and African ancestry, and γ captures the effects of the set of
1087 core covariates on gene expression.

1088 We reasoned that differences in the variance of gene expression between populations might
1089 inflate the number of false positives. We therefore used the *vcovHC* function of
1090 *sandwich* (v2.5-1) (ref.⁷¹) with the *type='HC3'* option to compute sandwich estimators of
1091 variance that are robust to residual heteroskedasticity. We estimated the β_r coefficients and
1092 their standard error with the *coefest* function of *lmtest* (v0.9-40) (ref.⁷²). FDR was calculated
1093 across all conditions and lineages with the Benjamini-Hochberg procedure (*p.adjust* function
1094 with '*fdr*' method). Genes with a FDR < 1% and $|\beta_r| > 0.2$ were considered to be as
1095 differentially expressed between populations (i.e., “raw” popDEGs). We adjusted for cellular
1096 composition within each lineage L , by introducing into model (3) a set of variables $(F_j)_{j \in L}$
1097 encoding the frequency in the PBMC fraction of each cell type j comprising the lineage (e.g.,
1098 naïve, effector, and regulatory subsets of CD4+ T cells).
1099

$$\text{Expr}_i = \alpha' + \beta_a \cdot I_i^{EUB} + Z_i^T \cdot \gamma' + \sum_{j \in L} \delta_j \cdot F_{j,i} + \varepsilon_i \quad (4)$$

1100 The notation is as above, with $\alpha', \beta_a, \gamma'$, the fitted parameters of the model. In this model, δ_j is
1101 the effect on gene expression of a 1% increase in cell type j and β_a indicates the cell
1102 composition-adjusted \log_2 fold change difference in expression between AFB and EUB
1103 individuals. The significance of β_a was calculated as described above, with a *sandwich*
1104 estimator of variance and the *coefest* function. FDR was calculated across all conditions and
1105 lineages to yield a set of “cell-composition-adjusted” popDEGs. We assessed the impact of
1106 cellular composition on differences in gene expression between populations, by defining
1107 Student’s test statistic $T_{\Delta\beta}$ as follows:

1108

$$T_{\Delta\beta} = \frac{\widehat{\beta}_a - \widehat{\beta}_r}{\text{Var}(\widehat{\beta}_a - \widehat{\beta}_r)} = \frac{\widehat{\beta}_a - \widehat{\beta}_r}{\widehat{s}_a^2 + \widehat{s}_r^2 - 2\rho\widehat{s}_a\widehat{s}_r} \quad (5)$$

1109

1110 where $\widehat{\beta}_r$ and $\widehat{\beta}_a$ are the raw and cell-composition-adjusted differences in expression
1111 between populations, \widehat{s}_r and \widehat{s}_a are the estimated standard error of $\widehat{\beta}_r$ and $\widehat{\beta}_a$ respectively,
1112 and ρ is the observed correlation in permuted data between the $\widehat{\beta}_r$ and $\widehat{\beta}_a$ statistics. Under the
1113 null hypothesis that population differences are not affected by cellular composition, $T_{\Delta\beta}$
1114 should follow an approximate Gaussian distribution with mean 0 and variance 1, enabling the
1115 definition of a p -value $p_{\Delta\beta}$. We then considered the set of raw popDEGs that (1) were not
1116 significant after adjustment (FDR > 1% or $|\beta_a| < 0.2$) and (2) displayed significant differences

1117 between the raw and adjusted effect sizes ($|T_{\Delta\beta}| > 1.96$) imputable to the effect of cellular
1118 composition.

1119 For the assessment of population differences in response to viral stimuli (i.e., popDRGs),
1120 we used the same approach, but with the replacement of log-normalized counts with the log-
1121 fold change difference in expression between the stimulation conditions for each of the two
1122 viruses and non-stimulated conditions.

1123

1124 **Pathway enrichment analyses**

1125 We performed functional assessments of the effects of cellular composition variability on
1126 differences in gene expression between donors in the basal state and in response to each virus,
1127 with the *fgsea* R package (v1.18.1) (ref.⁷³) and default options. This made it possible to
1128 perform a gene set enrichment analysis with population differences in each lineage ranked by
1129 the magnitude of the effect of ancestry on the expression or response of the gene before (β_r)
1130 and after (β_a) adjustment for differences in cellular composition.

1131

1132 **Fine mapping of expression quantitative trait loci (eQTL)**

1133 For eQTL mapping, we used variants with MAF >5% in at least one of the three populations
1134 considered, resulting in a set of 10,711,657 SNPs, of which 4,164,060 were located < 100kb
1135 from a gene. We used *MatrixEQTL* (v2.3) (ref.⁷⁴) to map eQTLs in a 100-kb region around
1136 each gene and obtain estimates of eQTL effect sizes and their standard error. eQTL mapping
1137 was performed separately for each immune lineage/cell type and condition, based on rank
1138 transformed gene expression values. eQTL analyses were performed adjusting for population,
1139 age, chromosomal sex, cell composition (within each lineage), as well as cell mortality and
1140 total number of cells in the sample, and a data-driven number of surrogate variables included
1141 to capture unknown confounders and remove unwanted variability. Specifically, for each
1142 immune lineage/cell type and condition, surrogate variables were obtained using the *sva*
1143 function from the *sva* R package (v3.40.0) (ref.⁷⁵) with option *method='two-steps'*, providing
1144 all other covariates as known confounders (*mod* argument). The number of surrogate
1145 variables to use in each lineage/cell type and condition was determined automatically based
1146 on the results from *num.sv* function with *method='be'* (ref.⁷⁵).

1147 For each gene, immune lineage/cell type and stimulation condition, Z-values (i.e., the
1148 effect size of each eQTL divided by the standard error of effect size) were then calculated,
1149 and the fine mapping of eQTLs was performed with *SuSiE* (v0.11.42) (ref.⁷⁶) (*susie_rss*
1150 function of the *susieR* R package), with a default value of up to 10 independent eQTLs per

1151 gene. Imputed genotype dosages were extracted in a 100-kb window around each gene and
1152 regressed against the population of origin (i.e., AFB, EUB or ASH). Genes with <50 SNPs in
1153 the selected window were discarded from the analysis. Pairwise correlations between the
1154 population-adjusted dosages were then assessed, to define the genotype correlation matrix to
1155 be used for the fine mapping of eQTLs. In rare cases (<0.1% of tested genes \times conditions
1156 combinations), the *susie_rss* function failed to converge, even when the number of iterations
1157 was increased to $>10^6$. These runs were, thus, discarded, and the associated eQTLs were
1158 assigned a null Z -score during FDR computation (see below). For each eQTL, the *index* SNP
1159 was defined as the SNP with the highest posterior inclusion probability (i.e., the α parameter
1160 in the output of *SuSiE*) for that eQTL, and the 95% credible interval was obtained as the
1161 minimal set of SNPs S such that $\alpha_s > 0.01$ for all $s \in S$ and $\sum_{s \in S} \alpha_s > 0.95$. Only eQTLs
1162 with a log-Bayes factor (*lbf*) > 3 were considered for further analyses.

1163 For each lineage and set of stimulation conditions, each eQTL identified by *SuSiE* was
1164 assigned an *eQTL evidence score*, defined as the absolute Z -value of association between the
1165 eQTL index SNP and the associated gene. We then used a pooled permutation strategy to
1166 define the genome-wide number of significant eQTLs (i.e., eQTL \times gene combinations)
1167 expected under the null hypothesis, for different thresholds of the *eQTL evidence score*. We
1168 repeated the eQTL mapping procedure on the dataset after randomly permuting genotype
1169 labels within each population. We then counted, for each possible *evidence score* threshold T ,
1170 the number of eQTLs identified in the observed and permuted data. Finally, we retained as a
1171 significance threshold the lowest threshold giving a number of significant eQTLs in the
1172 permuted data (false positives) of less than 1% the number of eQTLs identified in the
1173 observed data (false positives + true positives).

1174

1175 **Aggregation of eQTLs across cell types and stimulation conditions**

1176 The eQTL index SNP may differ between cellular states (immune lineage/cell type and
1177 stimulation condition), even in the presence of a single causal variant. It is therefore necessary
1178 to aggregate eQTLs to ensure that the same locus is tagged by a single variant across cellular
1179 states. To this end, we applied the following procedure, for each gene:

1180 1 Let C_g be the set of cellular states where a significant eQTL was detected for gene g ,
1181 and S_g be the associated list of eQTLs (i.e., cellular state \times index SNPs). We aim to define a
1182 minimal set of SNPs, M_g , that overlaps the 95% credible intervals of all significant eQTLs in
1183 S_g .

1184 2 For each SNP s in a 100-kb window around each gene, compute the expected number
1185 of cellular states where the SNP has a causal effect on gene expression $E[N_c(s)]$ as:

1186

$$1187 \quad E[N_c(s)] = \sum_{j \in C_g} PP_{sj} \quad (6)$$

1188

1189 where PP_{sj} is the posterior probability that SNP s has a causal effect on the expression of
1190 gene g in the cellular state j (cell type \times condition).

1191 3 Find the SNP s that maximizes $E[N_c(s)]$, and add it to M_g

1192 4 Remove from S_g , all eQTLs where the 95% credible interval contains SNP s

1193 5 Repeat steps 1-3 until S_g is empty.

1194

1195 At the end of this procedure, M_g provides the list of independent eQTL index SNPs (referred
1196 to as eSNPs) for gene g , for which we extracted summary statistics across all cellular states.

1197

1198 **Mapping of response eQTLs**

1199 For the mapping of response eQTLs (reQTLs), we repeated the same procedure as for the
1200 mapping of eQTLs, using rank-transformed \log_2 fold change as input rather than gene
1201 expression. This included reQTL mapping with *MatrixEQTL*⁷⁴, fine mapping with *SuSiE*⁷⁶,
1202 permutation-based FDR computation, and aggregation of reQTL across immune lineages, cell
1203 types and stimulation conditions. Surrogate variables were computed directly from \log_2 fold
1204 changes. For IAV-infected monocytes (detected only in the IAV condition), fold changes
1205 were computed relative to CD14⁺ monocytes in the non-stimulated condition. This produced a
1206 list of independent reQTL index SNPs M' , similar to that obtained for eQTLs, for which we
1207 extract summary statistics across all cellular states.

1208

1209 **Sharing of eQTLs across cell types and stimulation conditions**

1210 After extracting the set M of independent eSNPs across all genes, we defined ‘cell-type-
1211 specific eQTLs’ as eQTLs significant genome-wide in a single cell type. We accounted for
1212 the possibility that some eQTLs may be missed in specific cell types due to a lack of power,
1213 by introducing a second definition of eQTL sharing based on nominal p -values. Specifically,
1214 we considered an eQTL to be cell type-specific at a nominal significance if, and only if, it was
1215 significant genome-wide in a single cell type and its nominal p -value of association was
1216 greater than 0.01 in all other cell types. For each pair of cell types, the correlation of eQTL

1217 effect sizes was calculated on the set of all eQTLs passing the nominal significance criterion
1218 ($p < 0.01$) in at least one of the two cell types. To understand how the effect of genetics on
1219 immune response varies between SARS-CoV-2 and IAV, we defined an interaction statistic
1220 that enabled us to test for differences in reQTL effect size between the two viruses.

1221 Specifically, within each immune lineage/cell type, we defined:

1222

$$1223 \quad T_{int} = \frac{\widehat{\beta}_{IAV} - \widehat{\beta}_{COV}}{\text{Var}(\widehat{\beta}_{IAV} - \widehat{\beta}_{COV})} = \frac{\widehat{\beta}_{IAV} - \widehat{\beta}_{COV}}{s_{IAV}^2 + s_{COV}^2} \quad (7)$$

1224

1225 When the reQTL effect size is identical between the two viruses, we expect T_{int} to be
1226 normally distributed around 0 with variance 1, allowing to derive an interaction p -value. We
1227 thus defined as *virus-dependent reQTLs* those with a nominal interaction p -value < 0.01 and
1228 as *virus-specific reQTLs* those that passed a nominal p -value threshold of 0.01 in only one of
1229 the two stimulation conditions.

1230

1231 **Mediation analyses**

1232 For all popDEGs and popDRGs, we evaluated the proportion of the difference in expression
1233 or response to viral stimulation between populations attributable to either genetic factors (i.e.,
1234 eQTLs) or cellular composition, with the *mediate* function of the *mediation* package of R
1235 (v4.5.0) (ref.⁷⁷). Mediation analysis made it possible to separate the differences in
1236 expression/response between populations that were mediated by genetics (i.e., differences in
1237 allele frequency of a given eQTL between populations, ζ_g), or cellular composition (i.e.,
1238 difference in cell type proportions between populations, ζ_c) from those occurring
1239 independently of the eQTL/cell type considered (independent or direct effect δ). It was then
1240 possible to estimate the respective proportion of population differences mediated by genetics

$$1241 \quad \tau_g \text{ and cellular composition } \tau_c \text{ as } \tau_c = \frac{\zeta_c}{\zeta_c + \zeta_g + \delta} \text{ and } \tau_g = \frac{\zeta_g}{\zeta_g + \zeta_c + \delta}, \text{ with } \zeta_c + \zeta_g + \delta$$

1242 corresponding to the total differences in expression/response between populations. For each
1243 popDEG and popDRG, we focused on either (i) the most strongly associated SNP in a 100-kb
1244 window around the gene, regardless of the presence or absence of a significant (r)eQTL, or
1245 (ii) the cell type differing most strongly between populations in each lineage (i.e., CD16⁺
1246 monocytes in the myeloid lineage, κ light chain-expressing memory B cells in the B-cell
1247 lineage, effector cells in CD4⁺ T cell lineage, CD8⁺ EMRA cells in the CD8⁺ T-cell lineage,

1248 and memory cells in the NK cell lineage). For each popDEG and potential mediator M , we
1249 then ran *mediate* with the following models:

1250

$$1251 \quad Expr_i \sim \alpha + \beta \cdot M_i + \delta \cdot I_i^{EUB} + Z_i^T \cdot \gamma + \varepsilon_i \quad (8)$$

$$M_i \sim \alpha' + \delta' \cdot I_i^{EUB} + Z_i^T \cdot \gamma' + \varepsilon'_i \quad (9)$$

1252

1253 where $Expr_i$ corresponds to normalized expression values in the cell type/condition under
1254 consideration, α and α' are two intercepts, β is the effect of the mediator M_i on gene
1255 expression, δ and δ' are the (direct) effect of population (captured through the indicator
1256 variable I_i^{EUB}) on gene expression and on the mediator, γ and γ' capture the confounding
1257 effect of covariates (i.e. age, and cell mortality) on both gene expression and the mediator,
1258 and ε_i and ε'_i are normally distributed residuals. For popDRGs, we used the same approach,
1259 replacing normalized gene expression values with the \log_2 fold change in gene expression
1260 between the stimulated and unstimulated states.

1261

1262 **Detection of signals of natural selection**

1263 We avoided SNP ascertainment bias, by performing natural selection analyses with high-
1264 coverage sequencing data from the 1,000 Genomes (1KG) Project⁷⁸. We downloaded the
1265 GRCh38 phased genotype files from the New York Genome Center FTP server and
1266 calculated the pairwise F_{ST} (ref.⁷⁹) between our three study populations (AFB, EUB, or ASH)
1267 and all 1KG populations, to identify the 1KG populations who were the most genetically
1268 similar to our study populations. All selection and introgression analyses (see section
1269 ‘Archaic introgression analyses’) were based on the Yoruba from Ibadan, Nigeria (YRI), Utah
1270 residents with Northern and Western European ancestry (CEU) and Southern Han Chinese
1271 (CHS) populations, as these 1KG populations had the lowest F_{ST} values with our three study
1272 groups. We filtered the data to include only autosomal bi-allelic SNPs and insertions/deletions
1273 (indels), and removed sites that were invariant (i.e., monomorphic) across the three
1274 populations. We identified loci presenting signals of positive selection (local adaptation) with
1275 the population branch statistic (PBS)³⁸, based on the Reynold’s F_{ST} estimator⁷⁹ between pairs
1276 of populations. PBS values were calculated for the YRI, CEU, and CHS populations
1277 separately, with the other two populations used as the control and outgroup. For each
1278 population, genome-wide PBS values were then ranked, and variants with PBS values within
1279 the top 1% were considered putative targets of selection. For annotation of the selected

1280 eQTLs, the ancestral and derived states at each site were inferred from six-way EPO multiple
1281 alignments for primate species (EPO6, available from [ftp://ftp.ensembl.org/pub/release-](ftp://ftp.ensembl.org/pub/release-71/emf/ensembl-compara/epo_6_primate/)
1282 [71/emf/ensembl-compara/epo_6_primate/](ftp://ftp.ensembl.org/pub/release-71/emf/ensembl-compara/epo_6_primate/)), and the effect size was reported for the derived
1283 allele. For sites without an ancestral/derived state in the EPO6 alignment, the effect of the
1284 allele with the lowest frequency worldwide was reported.

1285 We assessed the extent to which different sets of eQTLs displayed signals of local
1286 adaptation in permutation-based enrichment analyses. For each population, we compared the
1287 mean PBS values at (r)eQTLs for each set of cell type/stimulation condition with the mean
1288 PBS values obtained for 10,000 sets of randomly resampled sites. Resampled sites were
1289 matched with eQTLs for minor allele frequency (mean MAF across the three populations,
1290 bins of 0.01), LD scores (quintiles), and distance to the nearest gene (bins of 0-1 kb, 1 kb-5
1291 kb, 10 kb-20 kb, 20 kb-50 kb, >100 kb). For each population and set of eQTLs, we defined
1292 the fold-enrichment (FE) in positive selection as the ratio of observed/expected values for
1293 mean PBS and extracted the mean and 95% confidence interval of this ratio across all
1294 resamplings. One-sided resampling p -values were calculated as the number of resamplings
1295 with a $FE > 1$ divided by the total number of resamplings. Resampling p -values were then
1296 adjusted for multiple testing by the Benjamini-Hochberg method.

1297

1298 **Detecting and dating episodes of local adaptation**

1299 We inferred the frequency trajectories of all eQTLs and reQTLs during the past 2,000
1300 generations (i.e., 56,000 years before the present, with a generation time of 28 years),
1301 systematically by using *CLUES* (commit no. 7371b86, 27 may 2021) (ref.³⁹). We first used
1302 *Relate* (v1.1.8) (ref.⁸⁰) on each population separately, to reconstruct tree-like ancestral
1303 recombination graphs (ARGs) around each SNP in the genome and to estimate effective
1304 population sizes across time based on the rate of coalescence events over the inferred ARGs.
1305 Using *CLUES*³⁹, we then estimated at each eQTL or reQTL, the most likely allele frequency
1306 trajectories for each sampled ARG and averaged these trajectories across all possible ARGs.

1307 We then analyzed changes in inferred allele frequencies over time, to identify selection
1308 events characterized by a rapid change in allele frequency. We considered the posterior mean
1309 of allele frequency at each generation and smoothed the inferred allele frequency trajectories
1310 by *loess* regression to ensure progressive changes in allele frequency over time, and to
1311 minimize the artifacts induced by the inference process. Finally, for each variant and in each
1312 population, we calculated the change in allele frequency f at each generation as the difference
1313 in the smoothed allele frequency between two consecutive generations:

$$\dot{f}(t) = \frac{df}{dt}(t) = f(t+1) - f(t) \quad (10)$$

1314 Under an assumption of neutrality, the count of a particular allele at generation $t+1$ is the
1315 result of a Bernoulli trial parameterized $B(N, f)$, where N is the size of the haploid population.
1316 The variance of allele frequencies at generation $t+1$ is, therefore, greater for alleles present at
1317 higher frequencies in generation t ,

$$V[f] = \frac{f(1-f)}{N} \quad (11)$$

1318 We adjusted for this, by scaling the change in allele frequency \dot{f} by a normalizing factor
1319 dependent on the allele frequency at generation t , such that the variance of the normalized
1320 change in allele frequency \dot{g} was the same across all variants,

$$\dot{g} = \frac{\dot{f}}{\sqrt{f(1-f)}} \quad (12)$$

1321 Finally, at each generation, we divided the normalized change in allele frequency \dot{g} by its
1322 standard deviation across all eQTLs and reQTLs, to calculate a Z -score for detecting alleles
1323 for which the normalized change in allele frequency exceeded genome-wide expectations,

$$Z = \frac{\dot{g}}{sd(\dot{g})} \quad (13)$$

1324 For each variant and generation, we then considered an absolute Z -score > 3 to constitute
1325 evidence of selection and we inferred the onset of selection of a variant as the first generation
1326 in which $|Z| > 3$.

1327

1328 **Simulations, power, and type I error estimates**

1329 We assessed the ability of our approach to detect (and date) events of natural selection
1330 correctly from the trajectories of allele frequencies, by using simulations with SLiM (v4.0.1)
1331 (ref.⁸¹) under various selection scenarios. Simulations were performed under a Wright-Fisher
1332 model for a single mutation occurring ~ 5000 generations ago, at a frequency varying from
1333 $f_{min} = \frac{1}{N}$ to $f_{max} = 1 - \frac{1}{N}$ in steps of 1%, where N is the simulated population size. We
1334 allowed population size to vary over time according to published estimates⁸⁰ for the YRI,
1335 CEU and CHS populations (Supplementary Fig. S9b). We then performed simulations both
1336 under an assumption of neutrality (1000 simulations for each starting frequency) and
1337 assuming a 200 generations-long episode of selection (100 simulations for each starting
1338 frequency and selection scenario). Selection episodes were simulated with an onset of
1339 selection 1000, 2000, 3000, or 4000 generations ago, and with a selection coefficient ranging

1340 from 0.01 to 0.05 (Supplementary Fig. S9c). We saved computation time, by performing a 10-
1341 fold scaling in line with SLiM recommendations. For each selected scenario and variant,
1342 simulated allele frequencies were retrieved every 10 generations, and smoothed with the *loess*
1343 function of R using default parameters. We then calculated normalized differences in
1344 smoothed allele frequencies for each simulated variant and scaled these differences at each
1345 generation, based on their standard deviation among neutral variants, to obtain *Z*-scores. For
1346 each selection scenario, we focused on the center of the selection interval and determined the
1347 type I error and power for various thresholds of absolute *Z*-scores varying from 0 to 6. We
1348 found that a threshold of 3 yielded both a low type I error (<0.2% false positives) and a
1349 satisfactory power for detecting selection events (Supplementary Fig. S9c). Finally, at each
1350 generation, we estimated the percentage of simulations, under an assumption of neutrality or a
1351 particular selection scenario, for which the absolute *Z*-score exceeded a threshold of 3. We
1352 found that significant *Z*-scores were equally rare at each generation under the assumption of
1353 neutrality, but that selected variants presented a clear and localized enrichment in significant
1354 *Z*-scores for intervals in which we simulated selection (Supplementary Fig. S9d).

1355

1356 **Archaic introgression analyses**

1357 For the definition of regions of the modern human genome of archaic ancestry (Neanderthal
1358 or Denisovan), we downloaded the VCFs from the high-coverage Neanderthal Vindija⁸² and
1359 Denisovan Altai⁸³ genomes (human genome assembly GRCh37;
1360 <http://cdna.eva.mpg.de/neandertal/Vindija/>) and applied the corresponding genome masks
1361 (FilterBed files). We then removed sites within segmental duplications and lifted over the
1362 genomic coordinates to the GRCh38 assembly with *CrossMap* (v0.6.3) (ref.⁸⁴). We used two
1363 statistics to identify introgressed regions in the CEU and CHS populations: (i) conditional
1364 random fields (CRF)^{85,86}, which uses reference archaic and outgroup genomes to identify
1365 introgressed haplotypes; and (ii) the S' method⁸⁷, which identifies stretches of probably
1366 introgressed alleles without requiring the definition of an archaic reference genome.

1367 For CRF-based calling, we phased the data with *SHAPEIT4* (v4.2.1) (ref.⁵⁵), using the
1368 recommended parameters for sequence data, and focused on bi-allelic SNPs for which the
1369 ancestral/derived state was unambiguously defined. We then performed two independent runs
1370 of CRF to detect haplotypes inherited from Neanderthal or Denisova. For Neanderthal-
1371 introgressed haplotypes, we used the Vindija Neanderthal genome as the archaic reference
1372 and YRI individuals merged with the Altai Denisovan genome as the outgroup. For
1373 Denisovan-introgressed haplotypes, we used the Altai Denisovan genome as the archaic

1374 reference panel and YRI individuals merged with the Vindija Neanderthal genome as the
1375 outgroup. Results from the two independent CRF runs were analyzed jointly, and we retained
1376 alleles with a marginal posterior probability $P_{\text{Neanderthal}} \geq 0.9$ and $P_{\text{Denisova}} < 0.5$ as
1377 Neanderthal-introgressed haplotypes and those containing alleles with $P_{\text{Denisova}} \geq 0.9$ and
1378 $P_{\text{Neanderthal}} < 0.5$ as Denisovan-introgressed haplotypes. For the S'-based calling of
1379 introgressed regions, we considered all biallelic SNPs with an allele frequency $< 1\%$ in the
1380 YRI population to be Eurasian-specific alleles. We then ran the *Sprime* software
1381 (v.07Dec18.5e2) (<https://github.com/browning-lab/sprime>) separately for the CEU and CHS
1382 populations, to identify and score putatively introgressed regions containing a high density of
1383 Eurasian-specific alleles. Putatively introgressed regions with a S' score $> 150,000$ were
1384 considered to be introgressed. This cutoff score has been shown to provide a good trade-off
1385 between power and accuracy based on simulations of introgression under realistic
1386 demographic scenarios⁸⁷. For both calling methods (i.e., CRF and S'), we used the
1387 recombination map from the 1,000 Genomes (1KG) Project Phase 3 data release⁵¹.

1388 After the calling of introgressed regions throughout the genome for each population, we
1389 defined SNPs of putative archaic origin (archaic SNPs, aSNPs) as those (i) located in an
1390 introgressed region defined by either the CRF or S' method, (ii) with one of their alleles being
1391 rare or absent ($\text{MAF} < 1\%$) in the YRI population, but present in the Vindija Neanderthal or
1392 Denisovan Altai genomes, and (iii) in high LD ($r^2 > 0.8$) with at least two other aSNPs and, to
1393 exclude incomplete lineage sorting, comprising an LD block of > 10 kb. This yielded a set of
1394 100,345 high-confidence aSNPs (Supplementary Table S7a). We further categorized aSNPs
1395 as of *Neanderthal origin*, *Denisovan origin* or *shared origin* according to their
1396 presence/absence in the Vindija Neanderthal and Denisovan Altai genomes. Finally, we
1397 considered any site that was in high LD with at least one aSNP in the same population in
1398 which introgression was detected to be *introgressed*, and classified introgressed haplotypes as
1399 of *Neanderthal origin*, *Denisovan origin*, or *shared origin* according to the most frequent
1400 origin of aSNPs in the haplotype. For introgressed SNPs, we defined the introgressed allele as
1401 (i) the allele rare or absent from individuals of African ancestry if the SNP was an aSNP, and
1402 (ii) for non-aSNPs, the allele most frequently segregating with the introgressed allele of
1403 linked aSNPs. In each population, introgressed alleles with a frequency in the top 1% for
1404 introgressed alleles genome-wide were considered to present evidence of *adaptive*
1405 *introgression*.

1406 The enrichment of introgressed haplotypes in eQTLs or reQTLs was assessed separately
1407 for each population (CEU and CHS), first by stimulation condition and then by cell type

1408 within each condition. Within each cell type/stimulation condition, we considered the set of
1409 all (r)eQTLs for which the index SNP displayed at least a marginal association (Student's
1410 $p < 0.01$) with gene expression. For each population and (r)eQTL set, we then grouped
1411 (r)eQTLs in high LD ($r^2 > 0.8$), retaining a single representative per group, and counted the
1412 total number of (r)eQTLs for which the index SNP was in LD ($r^2 > 0.8$) with an aSNP (i.e.,
1413 introgressed eQTLs). We then used *PLINK (v1.9) --indep-pairwise* (with a 500-kb window, 1
1414 kb step, an r^2 threshold of 0.8, and a MAF $> 5\%$) (ref.⁵²), to define tag-SNPs for each
1415 population, and we determined the expected number of introgressed SNPs by resampling tag-
1416 SNPs at random with the same distribution for MAF, LD scores, and distance to the nearest
1417 gene. We performed 10,000 resamplings for each (r)eQTL set and population. One-sided
1418 resampling-based p -values were calculated as the frequency at which the number of
1419 introgressed SNPs among resampled SNPs exceeded the number of introgressed SNPs among
1420 (r)eQTLs. Resampling-based p -values were then adjusted for multiple testing by the
1421 Benjamini–Hochberg method.

1422 We searched for signals of adaptive introgression, by determining whether introgressed
1423 haplotypes that altered gene expression were introgressed at a higher frequency than
1424 introgressed haplotypes with no effect on gene expression. For each stimulation cell
1425 type/condition, we focused on the set of introgressed eQTLs segregating with a MAF $> 5\%$ in
1426 each population (retaining a single representative per LD group) and compared the frequency
1427 of the introgressed allele with that of introgressed tag-SNPs genome-wide. We modeled
1428 $r_{(Freq)}$, the (rank-transformed) frequency of introgressed tag-SNPs according to the
1429 presence/absence of a linked eQTL (\mathbb{I}_{eQTL}), and the mean MAF of the SNP across the three
1430 populations (giving a higher power for eQTL detection).

1431

$$r_{(Freq)} \sim a + b \cdot \mathbb{I}_{eQTL} + c \cdot \overline{MAF} \quad (14)$$

1432

1433 where \mathbb{I}_{eQTL} is an indicator variable equal to 1 if the SNP is in LD with an eQTL ($r^2 > 0.8$)
1434 and 0 otherwise, \overline{MAF} is the mean MAF calculated separately for each population, a is the
1435 intercept of the model, b measures the difference in rank $r_{(Freq)}$ between eQTLs and non
1436 eQTLs, and c is a nuisance parameter capturing the relationship between \overline{MAF} and $r_{(Freq)}$.
1437 Under this model, the difference in frequency between eQTLs and non-eQTLs can be tested
1438 directly in a Student's t test with $H_0: b = 0$.

1439

1440 **Enrichment in COVID-19-associated loci and colocalization analyses**

1441 We downloaded summary statistics from the COVID-19 Host Genetics Initiative (release 7:
1442 <https://www.covid19hg.org/results/r7>) (ref.⁷) for three GWAS: (i) A2 - very severe respiratory
1443 cases of confirmed COVID-19 *vs.* the general population; (ii) B2 - hospitalized COVID-19
1444 cases *vs.* the general population; (iii) C2 – confirmed COVID-19 cases *vs.* the general
1445 population. We assessed the enrichment in eQTLs and reQTLs of COVID-19
1446 susceptibility/severity loci by considering, for each eQTL/reQTL, the A2, B2 or C2 *p*-values
1447 of the index SNP and calculating the percentage of eQTLs/reQTLs with a significant GWAS
1448 *p*-value of 10^{-4} . This percentage was then compared to that obtained for the resampled set of
1449 SNPs, matched for distance to the nearest gene (bins of 0-1, 1-5, 5-10, 10-20, 20-50, and 50-
1450 100 kb) and MAF (1% MAF bins). We performed 10,000 resamplings for each set of
1451 eQTLs/reQTLs tested. The use of different *p*-value thresholds for COVID-19-associated hits
1452 (10^{-3} to 10^{-5}) yielded similar results

1453 To identify specific eQTLs/reQTLs colocalized with GWAS hits, we first considered all
1454 (r)eQTLs for which the index SNPs were located within 100 kb of a SNP associated with
1455 COVID-19 susceptibility/severity (*p*-value $< 10^{-5}$). For each immune lineage/cell type, and
1456 condition for which the eQTL/reQTL reached genome-wide significance, we next extracted
1457 all SNPs in a 500-kb window around the index SNP for which summary statistics were
1458 available for both the eQTLs/reQTLs and COVID-19 GWAS phenotypes (A2, B2, and C2)
1459 and performed a colocalization test using the *coloc.signals* function of the *coloc* (v5.1.0)
1460 package of R. We set a prior probability for colocalization p_{12} of 10^{-5} (i.e., the recommended
1461 default value). Any pair of eQTL or reQTL/COVID-19 phenotypes with a posterior
1462 probability for colocalization $PP.H4 > 0.8$ was considered to display significant
1463 colocalization.

1464

1465 **Statistical Analyses**

1466 Unless explicitly specified, all statistical tests are two-sided and based on measurements from
1467 independent samples.

1468

1469

1470 **Data availability**

1471 The RNA sequencing data generated and analyzed in this study have been deposited in the
1472 Institut Pasteur data repository, OWEY, which can be accessed via the following link:

1473 <https://doi.org/XXXX>. The genome-wide genotyping data generated or used in this study
1474 have been deposited in OWEY and can be accessed at the following URL:

1475 <https://doi.org/XXXX>. Data access and use is restricted to academic research related to the
1476 variability of the human immune response.

1477

1478 **Code availability**

1479 All custom computer code or algorithms used in this study are available from github
1480 (https://github.com/h-e-g/popCell_SARS-CoV-2).

1481

1482 **Inclusion and Ethics**

1483 The current research project builds on samples collected in Ghent (Belgium) and Hong-Kong
1484 SAR (China) and has been conducted in collaboration with local researchers. Roles and
1485 responsibilities were agreed amongst collaborators ahead of the research. Research conducted
1486 in this study is relevant to local participants and has been reviewed by local ethics committees
1487 (committee of Ghent University, Belgium, n° B670201214647; Institutional Review Board of
1488 the University of Hong-Kong; n° UW 20-132), and the relevant French authorities (CPP,
1489 CCITRS and CNIL). This study was also monitored by the Ethics Board of Institut Pasteur
1490 (EVOIMMUNOPOP-281297). All manipulations of live viruses were performed in a high
1491 security BSL-3 environment.

1492

1493 **Methods references**

- 1494 51. 1000 Genomes Project Consortium *et al.* A global reference for human genetic
1495 variation. *Nature* **526**, 68-74 (2015).
- 1496 52. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and
1497 richer datasets. *Gigascience* **4**, 7 (2015).
- 1498 53. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**(2021).
- 1499 54. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for
1500 analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
- 1501 55. Delaneau, O., Zagury, J.F., Robinson, M.R., Marchini, J.L. & Dermitzakis, E.T.
1502 Accurate, scalable and integrative haplotype estimation. *Nat Commun* **10**, 5436
1503 (2019).
- 1504 56. Browning, B.L., Zhou, Y. & Browning, S.R. A One-Penny Imputed Genome from
1505 Next-Generation Reference Panels. *Am J Hum Genet* **103**, 338-348 (2018).
- 1506 57. Pepe, A., Pietropaoli, S., Vos, M., Barba-Spaeth, G. & Zurzolo, C. Tunneling
1507 nanotubes provide a route for SARS-CoV-2 spreading. *Sci Adv* **8**, eabo0171 (2022).
- 1508 58. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile
1509 mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv*
1510 <https://doi.org/10.1101/2021.05.05.442755> (2021).
- 1511 59. Kang, H.M. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural
1512 genetic variation. *Nat Biotechnol* **36**, 89-94 (2018).
- 1513 60. Amezquita, R.A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nat*
1514 *Methods* **17**, 137-145 (2020).
- 1515 61. Haghverdi, L., Lun, A.T.L., Morgan, M.D. & Marioni, J.C. Batch effects in single-cell
1516 RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat*
1517 *Biotechnol* **36**, 421-427 (2018).
- 1518 62. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
1519 Harmony. *Nat Methods* **16**, 1289-1296 (2019).
- 1520 63. Darnell, M.E. & Taylor, D.R. Evaluation of inactivation methods for severe acute
1521 respiratory syndrome coronavirus in noncellular blood products. *Transfusion* **46**,
1522 1770-7 (2006).
- 1523 64. Meyer, S. *et al.* AIRE-Deficient Patients Harbor Unique High-Affinity Disease-
1524 Ameliorating Autoantibodies. *Cell* **166**, 582-595 (2016).
- 1525 65. FlowJo™ Software, V. (ed. Becton, D.a.C.) (Ashland, OR, 2019).

- 1526 66. Buttner, M., Miao, Z., Wolf, F.A., Teichmann, S.A. & Theis, F.J. A test metric for
1527 assessing single-cell RNA-seq batch correction. *Nat Methods* **16**, 43-49 (2019).
- 1528 67. McCarthy, D.J., Campbell, K.R., Lun, A.T. & Wills, Q.F. Scater: pre-processing,
1529 quality control, normalization and visualization of single-cell RNA-seq data in R.
1530 *Bioinformatics* **33**, 1179-1186 (2017).
- 1531 68. Bates, D., Machler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models
1532 Using lme4. *J Stat Soft* **67**(2015).
- 1533 69. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587
1534 e29 (2021).
- 1535 70. Zuber, V. & Strimmer, K. High-Dimensional Regression and Variable Selection Using
1536 CAR Scores. *Stat. Appl. Genet. Mol. Biol.* **10**, 34 (2011).
- 1537 71. Zeileis, A., Koll, S. & Graham, N. Various Versatile Variances: An Object-Oriented
1538 Implementation of Clustered Covariances in R. *J Stat Soft* **95**(2020).
- 1539 72. Zeileis, A. & Hothorn, T. Diagnostic Checking in Regression Relationships. In R
1540 News. <https://CRAN.R-project.org/doc/Rnews/> 2, 3 (2002).
- 1541 73. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv*
1542 <https://doi.org/10.1101/060012> (2021).
- 1543 74. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations.
1544 *Bioinformatics* **28**, 1353-8 (2012).
- 1545 75. Leek, J.T. & Storey, J.D. Capturing heterogeneity in gene expression studies by
1546 surrogate variable analysis. *PLoS Genet* **3**, 1724-35 (2007).
- 1547 76. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to
1548 variable selection in regression, with application to genetic fine-mapping. *bioRxiv*
1549 <https://doi.org/10.1101/501114> (2020).
- 1550 77. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R Package for
1551 Causal Mediation Analysis *J Stat Soft* **59**(2014).
- 1552 78. Byrska-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded
1553 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-3440 e19 (2022).
- 1554 79. Reynolds, J., Weir, B.S. & Cockerham, C.C. Estimation of the coancestry coefficient:
1555 basis for a short-term genetic distance. *Genetics* **105**, 767-79 (1983).
- 1556 80. Speidel, L., Forest, M., Shi, S. & Myers, S.R. A method for genome-wide genealogy
1557 estimation for thousands of samples. *Nat Genet* **51**, 1321-1329 (2019).
- 1558 81. Haller, B.C. & Messer, P.W. SLiM 3: Forward Genetic Simulations Beyond the
1559 Wright-Fisher Model. *Mol Biol Evol* **36**, 632-637 (2019).

- 1560 82. Prufer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia.
1561 *Science* **358**, 655-658 (2017).
- 1562 83. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan
1563 individual. *Science* **338**, 222-6 (2012).
- 1564 84. Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between genome
1565 assemblies. *Bioinformatics* **30**, 1006-7 (2014).
- 1566 85. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day
1567 humans. *Nature* **507**, 354-7 (2014).
- 1568 86. Sankararaman, S., Mallick, S., Patterson, N. & Reich, D. The Combined Landscape of
1569 Denisovan and Neanderthal Ancestry in Present-Day Humans. *Curr Biol* **26**, 1241-7
1570 (2016).
- 1571 87. Browning, S.R., Browning, B.L., Zhou, Y., Tucci, S. & Akey, J.M. Analysis of
1572 Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell*
1573 **173**, 53-61 e9 (2018).
- 1574

1575 **Acknowledgments**

1576 We thank all members of the Human Evolutionary Genetics Laboratory, and Bjorn-Axel Olin
1577 in particular, and Monique Van der Wijst, Dylan de Vries and Lude Franke for helpful
1578 discussions, Leo Speidel for help in running *Relate* and interpreting the results, and Anja
1579 Coen and Frederic Clément for assistance with sample collection. The Human Evolutionary
1580 Genetics Laboratory is supported by the *Institut Pasteur*, the *Collège de France*, the *Centre*
1581 *Nationale de la Recherche Scientifique* (CNRS), the *Agence Nationale de la Recherche*
1582 (ANR) grants COVID-19-POPCELL (ANR-21-CO14-0003-01), POPCELL-REG (ANR-22-
1583 CE12-0030-01) and COVIFERON (ANR-21-RHUS-0008), the EU HORIZON-HLTH-2021-
1584 DISEASE-04-07 grant UNDINE (no. 101057100), the French Government’s *Investissement*
1585 *d’Avenir* program, *Laboratoires d’Excellence* “Integrative Biology of Emerging Infectious
1586 Diseases” (ANR-10- LABX-62-IBEID) and “*Milieu Intérieur*” (ANR-10-LABX-69-01), the
1587 *Fondation pour la Recherche Médicale* (Equipe FRM DEQ20180339214), the *Fondation*
1588 *Allianz-Institut de France*, and the *Fondation de France* (no. 00106080). K.L., J.T.K.W. and
1589 M.P. are supported by the Health and Medical Research Fund Commissioned Research on the
1590 Novel Coronavirus Disease, Hong Kong SAR (COVID190126), S.A.V by the Theme-based
1591 Research Scheme of the Research Grants Council of the Hong Kong SAR (T11-705/21-N,
1592 SAV: T11-712/19-N), and M.P., R.B. and D.D. by *InnoHK*, an initiative of the Innovation
1593 and Technology Commission, the Government of the Hong Kong SAR.

1594

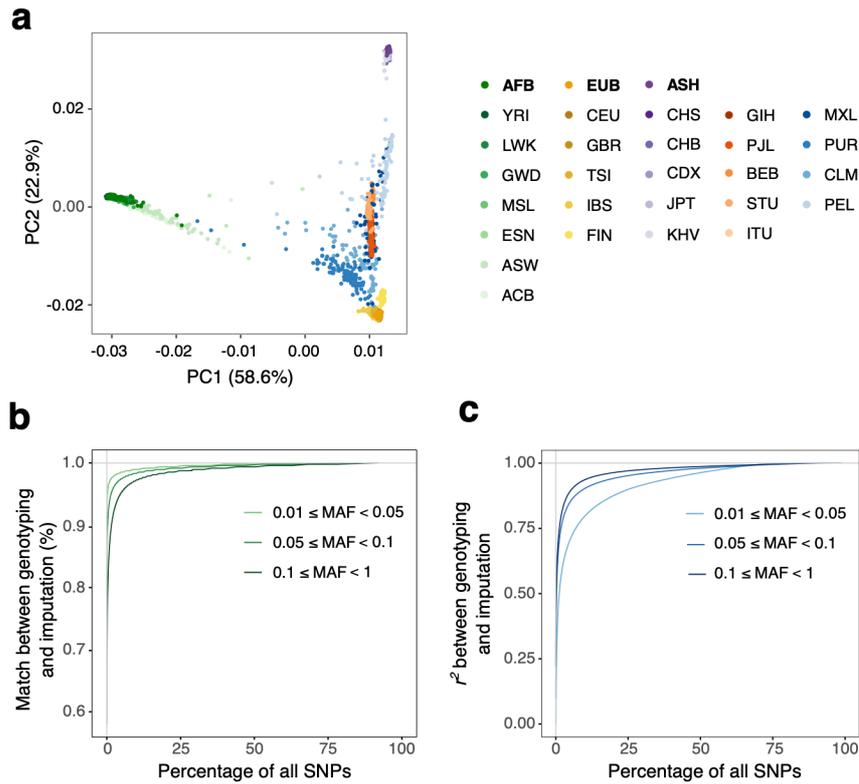
1595 **Author Contributions**

1596 M.R. and L.Q.-M. conceived and supervised the study; Y.A., A.B., M.O’N. and M.R.
1597 designed experiments; Y.A., A.B., Z.L., M.O’N. and S.H.M. conducted the experiments;
1598 Y.A., M.O’N, J.M.-R., G.K. and M.R. designed and performed computational analyses; V.B.
1599 conducted the SIMOA experiments; G.L.-R., C.-K.L., K.L., J.T.K.W., M.P., R.B. and S.A.V.
1600 managed clinical protocols and recruited donors; N.S., G.B.-S., S.P. and O.S. obtained SARS-
1601 CoV-2 strains and helped to design the stimulation experiments; C.C., M. M., M.H., V.L.,
1602 F.L, R.P.-R., L.A., J-L.C., D.D and E.P. advised on experiments and data analyses and
1603 interpretation; Y.A., A.B., M.R. and L.Q.-M. interpreted the data and wrote the manuscript,
1604 with critical input from all authors.

1605

1606 **Competing interests**

1607 None of the authors has any competing interests to declare.



1608

1609

1610 **Supplementary Figure 1 | Genetic structure of study populations and SNP imputation. a,**

1611 Principal component analysis of genotyping data. Each dot corresponds to an individual.

1612 Study samples (AFB, EUB, and ASH, in bold typeface) are projected jointly with 1,000

1613 Genomes populations of various ancestries including African (dots colored in green

1614 gradients), European (dots colored in yellow gradient), East Asian (dots colored in purple

1615 gradient), South Asian (dots colored in orange gradient) and American (dots colored in blue

1616 gradients). Abbreviations for each individual population can be found in ref.⁵¹. **b** and **c**,

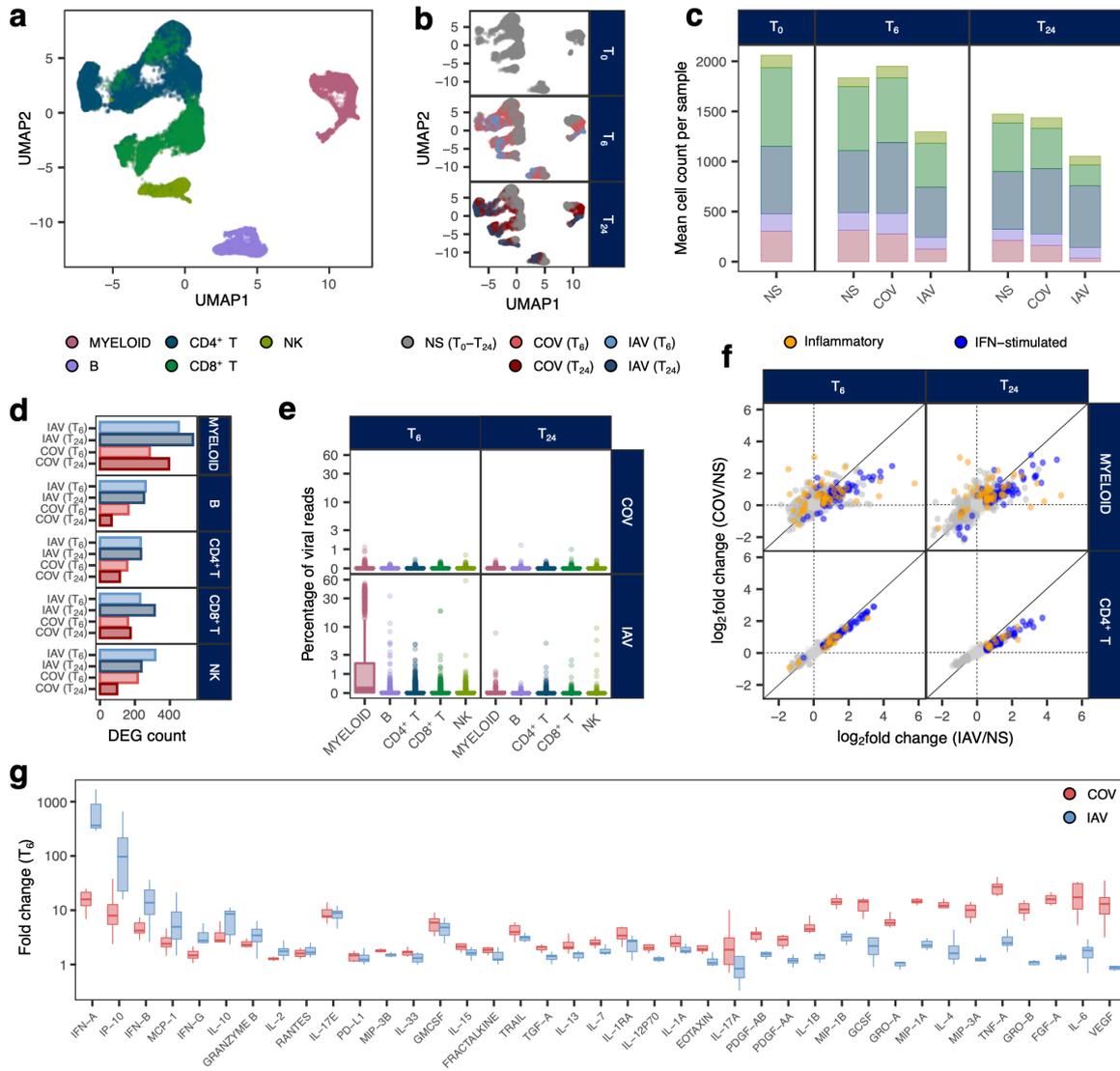
1617 Quality control of genotype imputation. Distribution of r^2 between genotyped and imputed

1618 SNPs (**b**) and genotype accuracy (**c**) obtained by 100-fold cross-validation and shown

1619 separately for different bins of minor allele frequency (MAF).

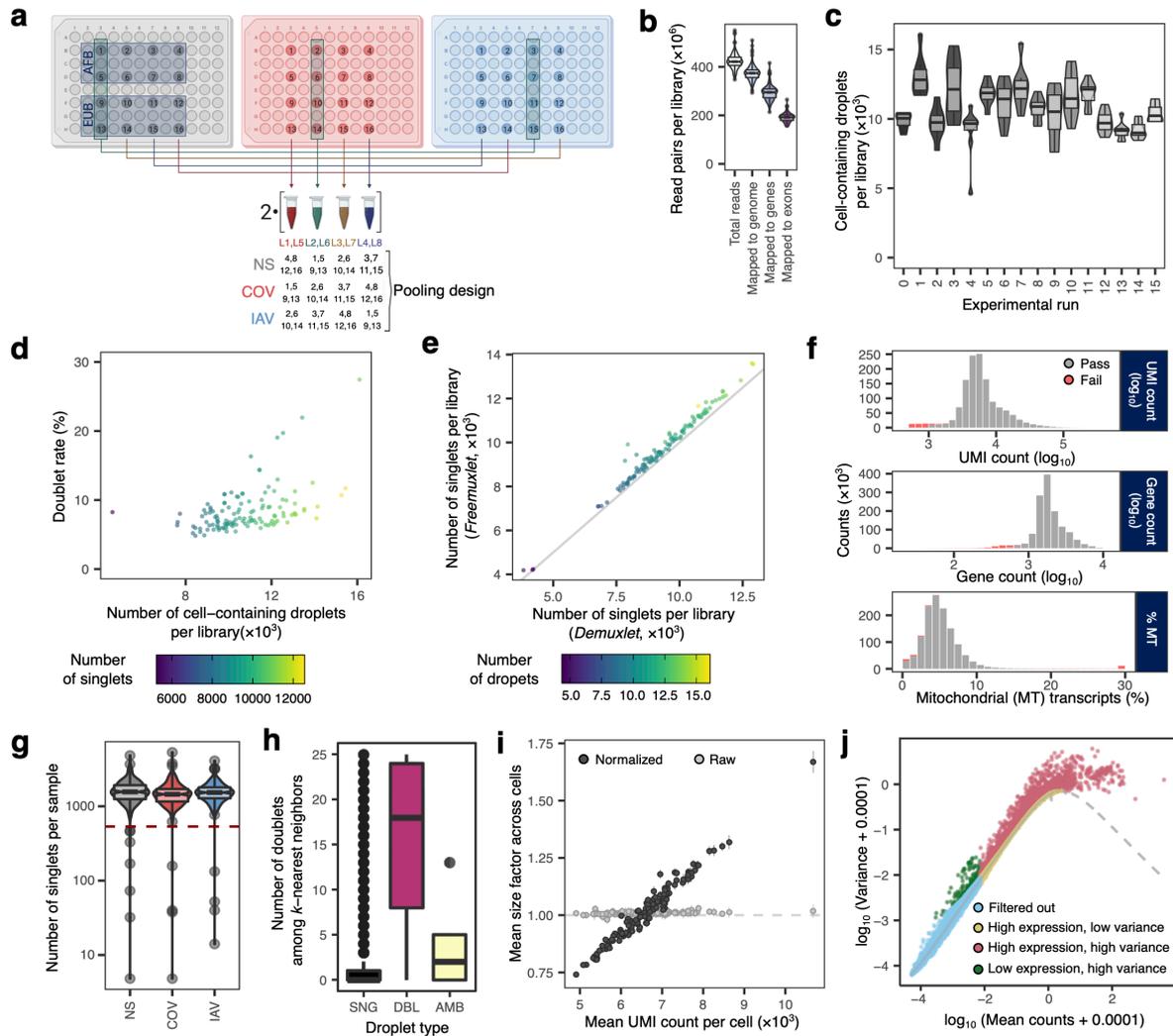
1620

1621



1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641

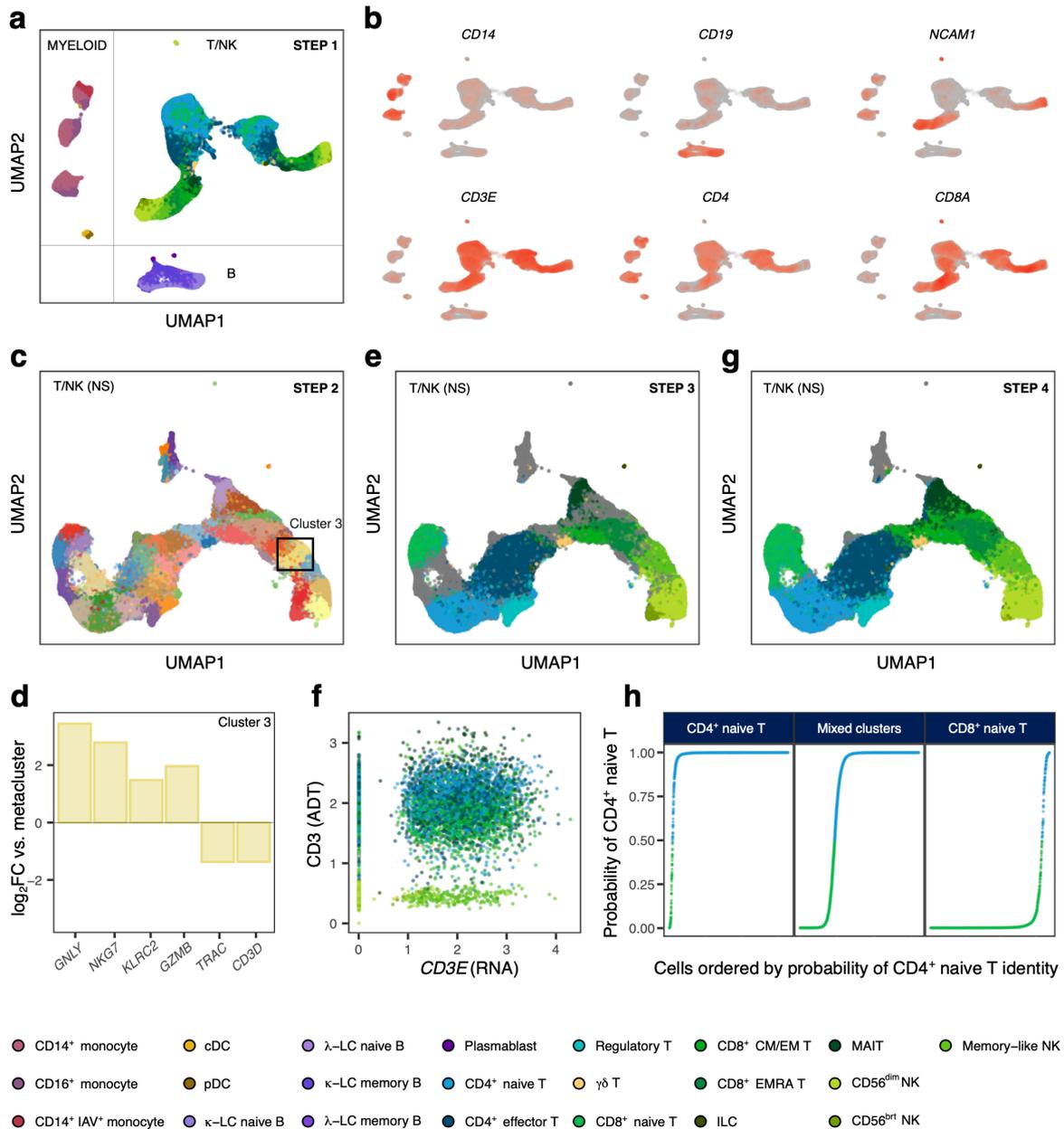
Supplementary Figure 2 | Single-cell kinetics of the immune response to RNA viruses. a, b, Uniform manifold approximation and projection (UMAP) of 86,363 peripheral blood mononuclear cells (PBMCs), mock-stimulated (NS) or stimulated with SARS-CoV-2 (COV) or influenza A virus (IAV) for 0, 6 or 24 hours. **c,** Mean cell-type counts per individual, set of stimulation conditions and time point. **d,** Number of differentially expressed genes (DEG; absolute log₂ fold change ($|\log_2FC|$) > 0.5, FDR < 0.01) after 6 or 24 hours of stimulation with SARS-CoV-2 or IAV relative to non-stimulated controls. **e,** Percentage of reads mapping to the SARS-CoV-2 or IAV genomes per cell, after 6 or 24 hours of stimulation, split by major immune lineage. **f,** Comparison of inflammatory and interferon-stimulated transcriptional responses of myeloid cells and CD4⁺ T cells (as an example of a lymphoid cell type) after 6 or 24 hours of stimulation with SARS-CoV-2 or IAV. **g,** Cytokine protein responses to 6 hours of stimulation with SARS-CoV-2 or IAV. In **a, c** and **e,** the colors indicate the immune lineages inferred from single-cell transcriptome data. In **b, d** and **g,** the colors indicate the stimulation condition and time post-stimulation. In **e** and **g,** boxplots are defined as follows: middle line, median; box limits, upper and lower quartiles; whiskers, 1.5 \times interquartile range; points, outliers.



1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663

Supplementary Figure 3 | Quality control of single-cell RNA-seq data. **a**, Experimental design. During each experimental run, PBMCs from 16 individuals (numbered 1 to 16) were processed in three different sets of experimental conditions (colored plates), and the resulting samples were split into four pools of 12 samples (four non-stimulated (NS), four influenza A virus-stimulated (IAV), four SARS-CoV-2-stimulated (COV)). Each pool was then processed on two independent libraries to increase the number of cells per sample (eight pools of 12 samples in total). **b**, Library sequencing depth, and total number of reads aligned either genome-wide, in the genic region, or over the coding exons. **c**, Distribution of the number of cell-containing droplets detected per library across the 16 experimental runs performed. **d**, Percentage of doublets per library as a function of the number of cell-containing droplets detected. Colors reflect the inferred number of singlets in the library. **e**, Number of singlets per library inferred with two independent demultiplexing algorithms: *Demuxlet* (supervised) and *Freemuxlet* (unsupervised). Colors reflect the total number of droplets in the library. **f**, Distribution of cells along quality-control metrics in our data set (i.e., UMI count, gene count, and percentage of mitochondrial reads). Cells that were excluded by our hard-threshold filtering are highlighted in red. **g**, Number of high-quality cells per sample (individual \times condition), split by stimulation condition. Individuals with <500 cells in at least one sample were excluded (dashed red line). **h**, Number of genetic doublets among k nearest neighbors as a function of the droplet status assigned by *Demuxlet* (SNG: singlet, DBL: doublet, AMB: ambiguous). **i**, Per-library mean of raw and batch-normalized size factors, as a function of the

1664 mean number of UMIs per cell in the library. Vertical bars indicate 95% CI of the mean. After
1665 normalization with *multiBatchNorm*, size factors successfully capture differences in read
1666 depth across libraries. **j**, Filtering of weakly expressed and low-variability genes. For each
1667 gene, the variance and mean counts are shown on a log scale (with an offset of 10^{-4}). The
1668 dashed line indicates the expected relationship between mean and variance under a Poisson
1669 distribution (technical noise). Genes are colored according to their expression levels (highly
1670 expressed: mean > 0.01) and estimated biological variance (highly variable: biological
1671 variance > 0.001). Genes that are both weakly expressed and of low variability (light blue) are
1672 excluded from downstream analyses. In **b**, **c**, **g**, and **h**, boxplots are defined as follows: middle
1673 line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; points,
1674 outliers.
1675



1676

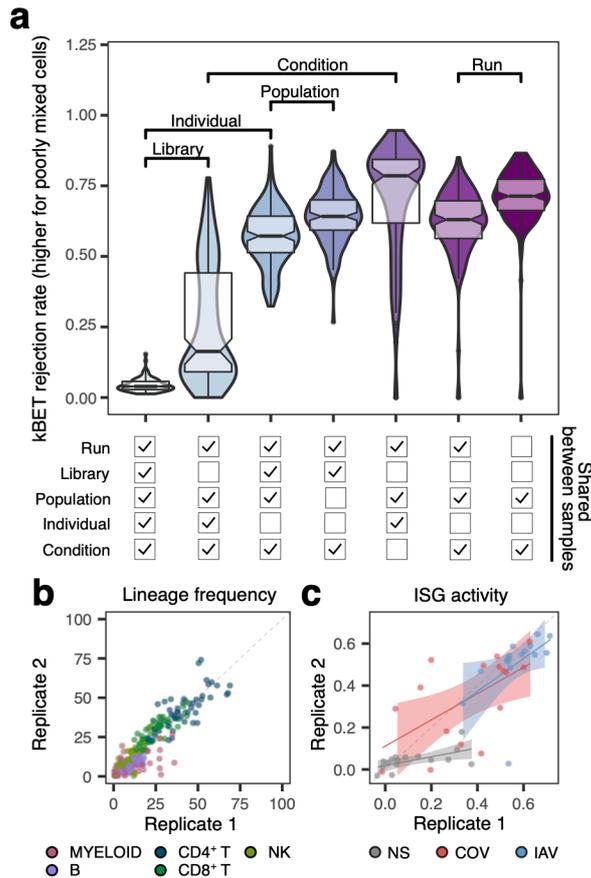
1677

1678

Supplementary Figure 4 | Cell-type assignment according to cluster-based transcriptional profiles and surface protein markers. **a**, Uniform manifold approximation and projection (UMAP) of 1,047,824 PBMCs, either resting (mock-stimulated) or stimulated with SARS-CoV-2 (COV) or influenza A virus (IAV) for 6 hours. **b**, Normalized single-cell RNA UMI count distributions of 6 canonical marker genes. **c**, Graph-based subclustering of the non-stimulated T/NK meta-cluster; cluster 3, initially defined as part of a larger cluster of mixed NK and CD8⁺ T cells, is highlighted. **d**, Log₂-fold change difference in expression of markers defining NK cell identity between cluster 3 and the rest of the T/NK metacluster. **e**, Cell-type inference based on canonical marker expression in sub-clusters increases the resolution of cell-type identities, but some mixed-identity and unidentified clusters remain (gray). **f**, At the transcriptional level, most of the cells in cluster 3 are CD3E-positive, and, thus, associated with lymphocyte lineages, but CITE-seq data show that most cluster 3 cells do not express CD3 protein, hence their assignment to the NK lineage. **g**, Cell-type inference after CITE-seq-based assignment and resolution of mixed-identity clusters by linear

1691

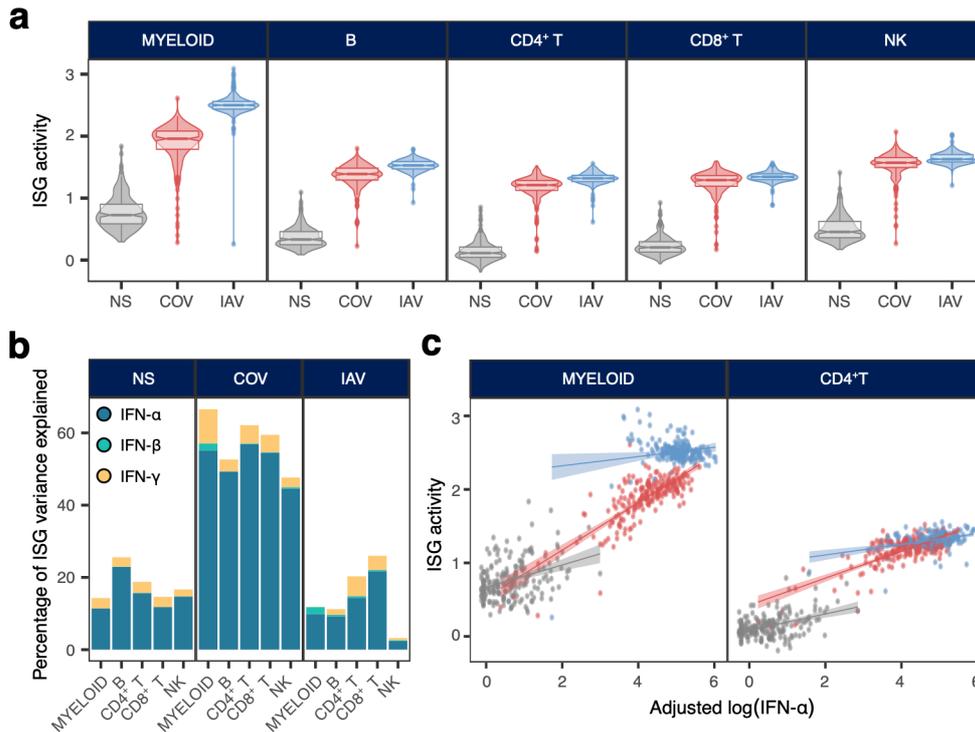
1692 discriminant analysis (LDA). Unassigned cells (gray) are discarded. **h**, Assignment of cells
1693 from mixed-identity clusters, based on previously identified clusters. In this example, LDA
1694 models are trained on data from 10,000 confidently identified naive CD4⁺ and naive CD8⁺ T
1695 cells, making it possible to assign most cells from a mixed cluster to one of the two target
1696 identities.



1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714

Supplementary Figure 5 | Batch effects and replicability of single-cell experiments. a,

Effect of technical and biological variation on the clustering of cells. For each comparison, we consider a random subset of 150 pairs of samples (run × library × individual × condition). For each pair of samples, violins and boxplots show the distribution of the kBET rejection rate, which increases when cells from different samples tend to cluster separately (middle line: median; box limits: upper and lower quartiles; whiskers: 1.5× interquartile range; points: outliers). For self-comparisons, cells from the same sample were randomly split into two groups before kBET calculation. Comparisons for quantifying the effects of various factors (e.g., run, library preparation, individual, population or stimulation condition) on cell mixing are highlighted. For all comparisons shown, differences in kBET are significant for a Wilcoxon rank-sum test p -value $< 10^{-9}$. **b** and **c**, Comparison of estimated lineage proportions (b) and mean ISG expression (c) between replicate samples processed separately across different runs (cells from the same individual thawed on a different day and stimulated in the same conditions). In **c**, regression lines and the 95% confidence interval are shown for each stimulation condition.



1715

1716

1717

1718

1719

1720

1721

1722

1723

1724

1725

1726

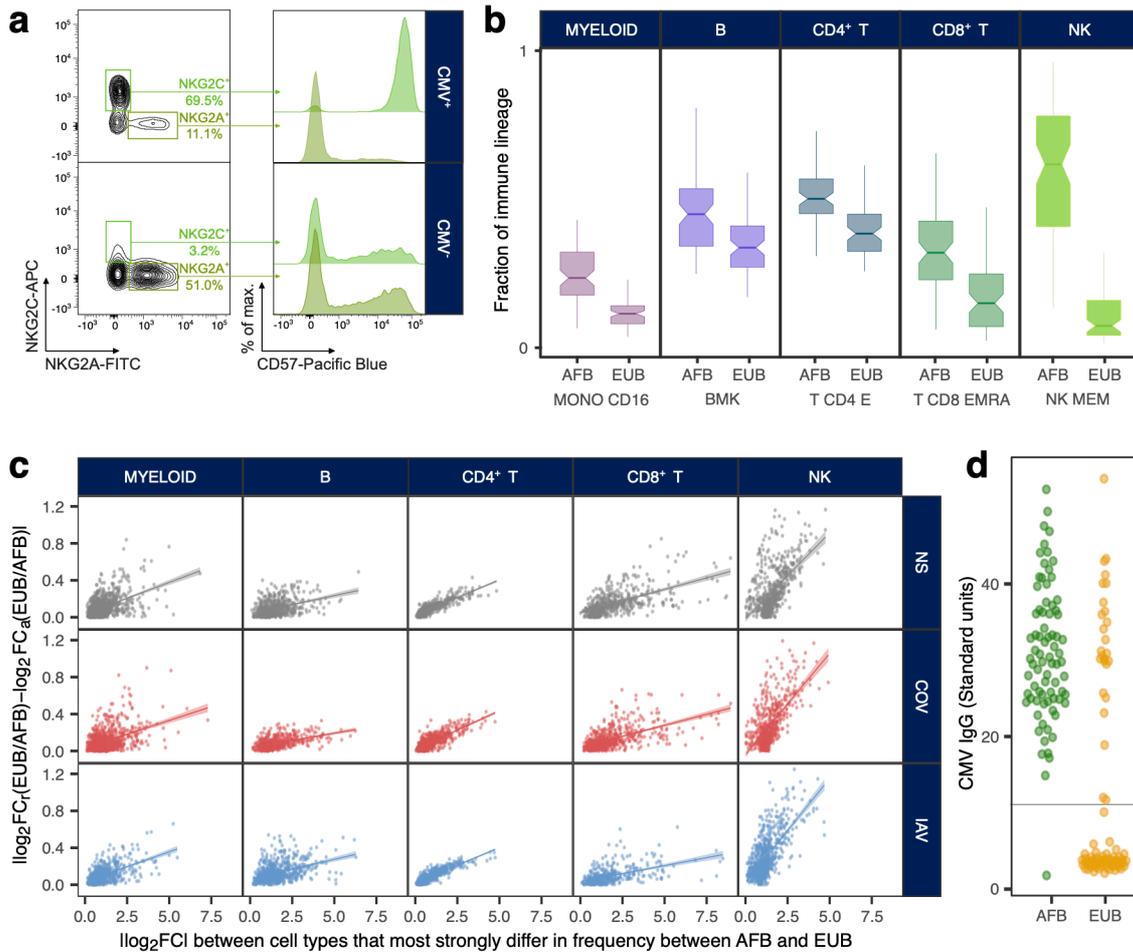
1727

1728

1729

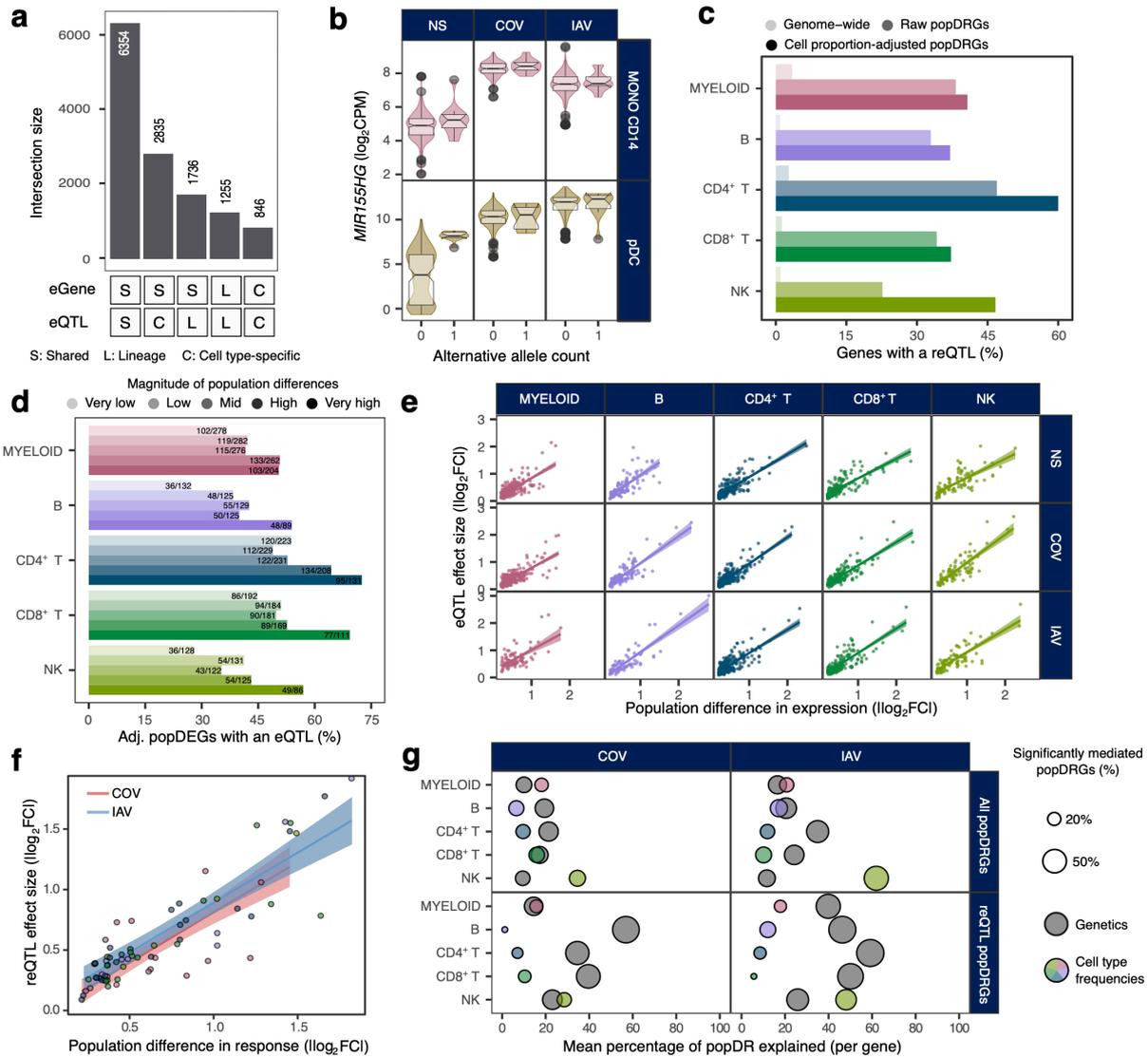
1730

Supplementary Figure 6 | Interferon-stimulated gene responses to SARS-CoV-2 and IAV stimulation. **a**, Distribution of ISG activity in the non-stimulated state and in response to SARS-CoV-2 (COV) and influenza A virus (IAV) across the five major immune lineages. For each lineage and set of stimulation conditions, the violins and boxplots show the distribution of ISG activity scores across all 222 donors (middle line: median; box limits: upper and lower quartiles; whiskers: 1.5 \times interquartile range; points: outliers). **b**, Proportion of the variance of ISG activity explained by IFN- α , IFN- β and IFN- γ in the non-stimulated condition and in response to SARS-CoV-2 and IAV, across the five major immune lineages. **c**, Correlation between levels of IFN- α in the supernatants (measured by SIMOA) and ISG activity in myeloid and lymphoid (CD4⁺ T cells) cells, adjusted for cellular mortality. Each dot represents a sample (donor \times condition) and is colored according to its stimulation condition (gray: NS, red: COV and blue: IAV). For each lineage and set of stimulation conditions, regression lines and 95% confidence intervals are shown.



1731
1732

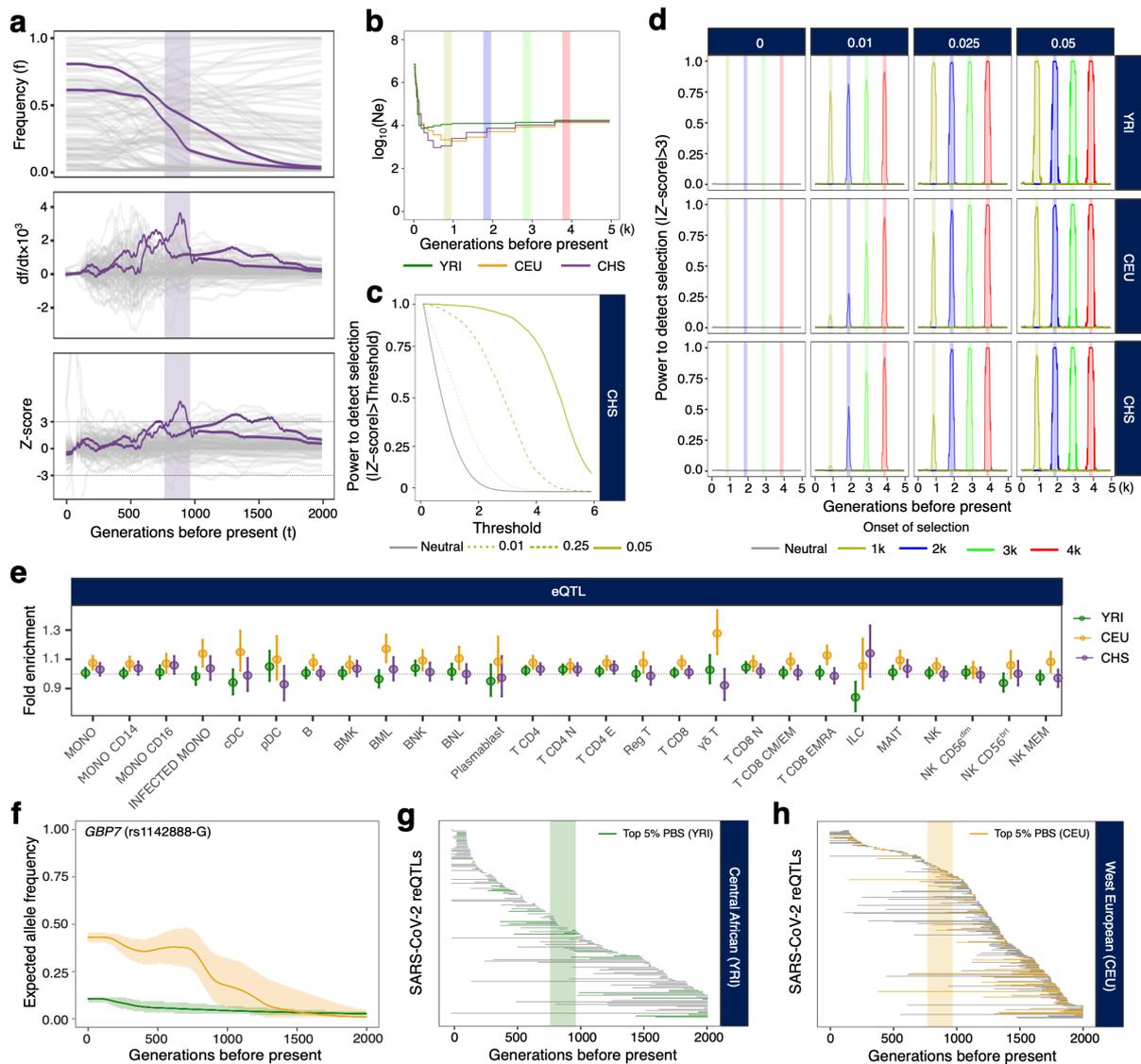
1733 **Supplementary Figure 7 | Population cellular heterogeneity and transcriptional**
 1734 **responses to viral stimulation.** **a**, Validation of the memory-like NK fraction. Flow
 1735 cytometry data for representative CMV⁺ and CMV⁻ donors, highlighting the higher
 1736 percentage of memory-like NK cells (NKG2C⁺, NKG2A⁻, CD57⁺) among CMV⁺ donors than
 1737 among CMV⁻ donors. **b**, Population variation in the percentage of CD16⁺ monocytes, memory
 1738 lymphocyte subsets and memory-like NK cells. For each major immune lineage, the cell type
 1739 differing most strongly in frequency between AFB and EUB donors is shown. Boxplots show
 1740 the distribution of the percentage of the target cell type in the corresponding lineage in each
 1741 population (middle line: median; box limits: upper and lower quartiles; whiskers: 1.5×
 1742 interquartile range). **c**, Effect of adjusting for cellular composition on the absolute differences
 1743 in expression between AFB and EUB donors, as a function of absolute differences in
 1744 expression between the two cell types differing most in frequency between these populations
 1745 (Supplementary Table 3b). For each lineage and stimulation condition, the regression line and
 1746 95% confidence interval are shown. **d**, Serology assays for CMV across donors according to
 1747 ancestry. Each dot represents a donor and is colored according to ancestry (AFB: Central
 1748 Africans, EUB: West Europeans). The gray line represents the detection threshold used to
 1749 identify a donor as seropositive.
 1750



1751
1752

1753 **Supplementary Figure 8 | Expression quantitative trait loci mapping and contribution to**
 1754 **population differences in response to RNA viruses.** **a**, Overlap of eQTLs and eGenes (i.e.,
 1755 genes with an eQTL) detected during the mapping of eQTLs at the immune lineage and cell-
 1756 type levels. **b**, Example of a pDC-specific eQTL for *MIR155HG*. *MIR155HG* expression
 1757 levels in pDCs and CD14⁺ monocytes according to rs114273142 genotype in non-stimulated
 1758 (NS), SARS-CoV-2-stimulated (COV) and influenza A virus-stimulated (IAV) conditions
 1759 (middle line: median; box limits: upper and lower quartiles; whiskers: 1.5× interquartile
 1760 range; points: outliers). **c**, Enrichment in reQTLs among popDRGs. For each lineage, bars
 1761 indicate the percentage of genes with a significant reQTL, both genome-wide and among the
 1762 popDRGs identified, before or after adjustment for cell composition (referred to as “adjusted”
 1763 and “raw” respectively). **d**, Percentage of popDEGs with an eQTL according to the magnitude
 1764 of differences in expression. In each lineage, popDEGs are assigned to one of five magnitude
 1765 groups based on quintiles of log₂fold change between the AFB and EUB populations. For
 1766 each lineage and magnitude group, the number of popDEGs with an eQTL and the total
 1767 number of popDEGs are reported. **e**, Relationship between eQTL effect sizes and population
 1768 differences in expression. **f**, Relationship between reQTL effect sizes and population
 1769 differences in response to immune stimulation. For each stimulation condition, the regression
 1770 line is computed jointly across all cell types. **g**, Contribution of genetics and cell composition
 1771 to population differences in response to stimulation by COV and IAV. For each lineage and

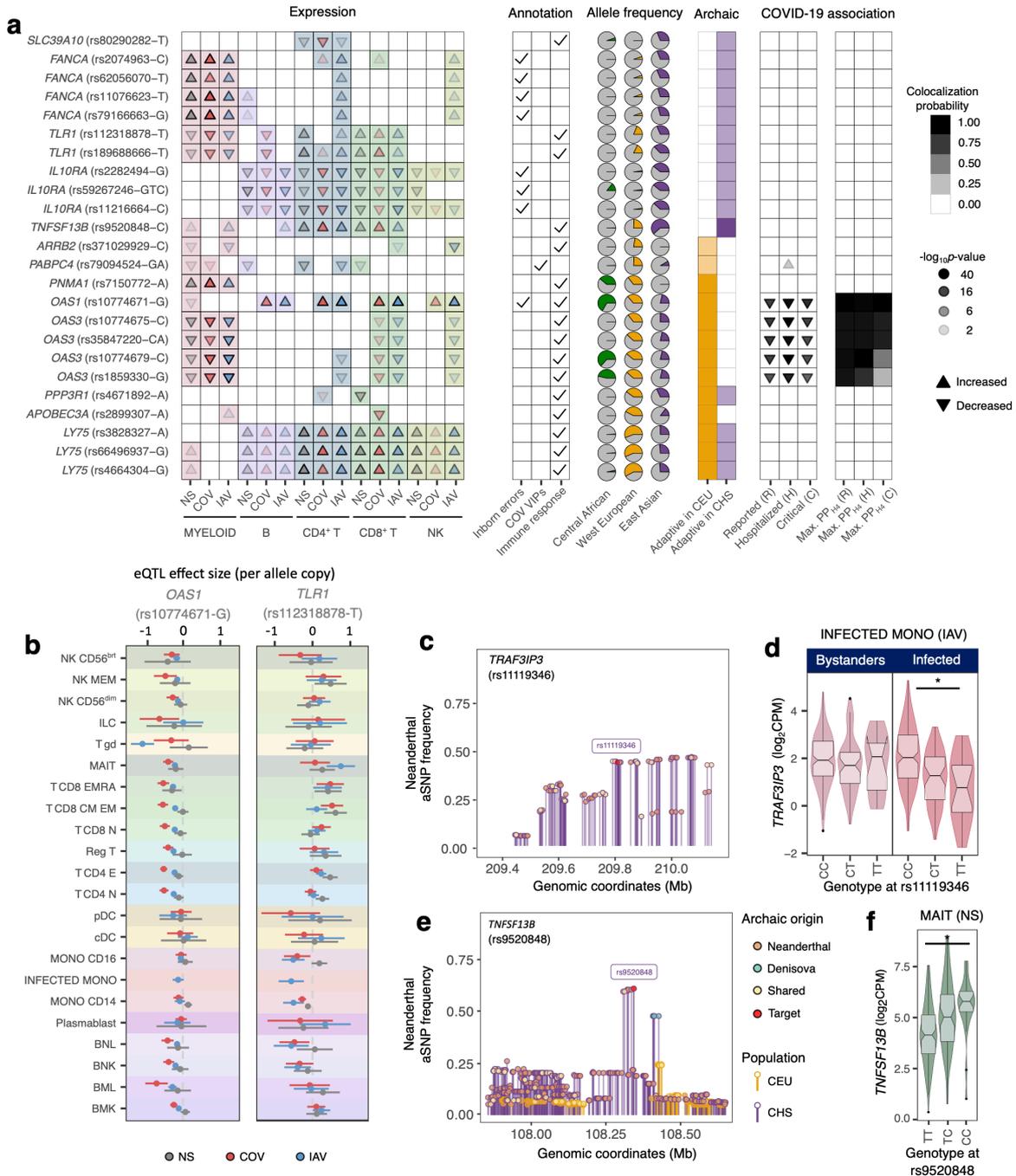
1772 stimulation condition, the x -axis indicates the mean percentage of population differences in
1773 response to stimulation mediated by either genetics or cell composition, across all popDRGs
1774 (upper panels) or the set of popDRGs with a significant reQTL (lower panels). The size of the
1775 dots reflects the percentage of genes with a significant mediated effect (FDR<1%).
1776



1777
 1778
 1779
 1780
 1781
 1782
 1783
 1784
 1785
 1786
 1787
 1788
 1789
 1790
 1791
 1792
 1793
 1794
 1795
 1796
 1797

Supplementary Figure 9 | Positive selection signals across time, cell types and populations. **a**, Method for estimating the time of onset of selection from derivative information from allele frequency trajectories. (Top) Allele frequency trajectories in an East Asian population (CHS) across the past 2,000 generations of two SARS-CoV-2 reQTLs (i.e., rs4806787 and rs1028396), affecting the response of *LILRB1* in plasmacytoid dendritic cells and *SIRPA* in CD14⁺ monocytes, respectively. (Middle) Change at each generation (from past to present) of the (smoothed) frequency of the derived allele, normalized for allele frequency. (Bottom) Z-score calculated as the normalized derivative, scaled at each generation by the standard deviation of derivatives across all eQTLs. Periods of selection are estimated as the range, in generations, over which the rate of change in the frequency of each allele deviates significantly from expectations under the hypothesis of neutrality (i.e., $|Z\text{-score}| > 3$). (Top to bottom) The corresponding allele frequency trajectories, first derivatives and Z-scores for 100 random SNPs sampled from the set of all (r)eQTLs detected in this study are shown in gray. **b**, Effective population size and episodes of positive selection over time used in our simulations. Colored lines indicate effective population size (green: YRI, yellow: CEU, purple: CHS); shaded areas indicate positive selection events. **c**, Type I error and power as a function of Z-score threshold. Power and type I error are reported for selection occurring 1000-1200 generations ago, for coefficients of selection ranging from 0.01 to 0.05. **d**, Power to detect positive selection at a Z-score threshold of 3, as a function of time. Lines are colored

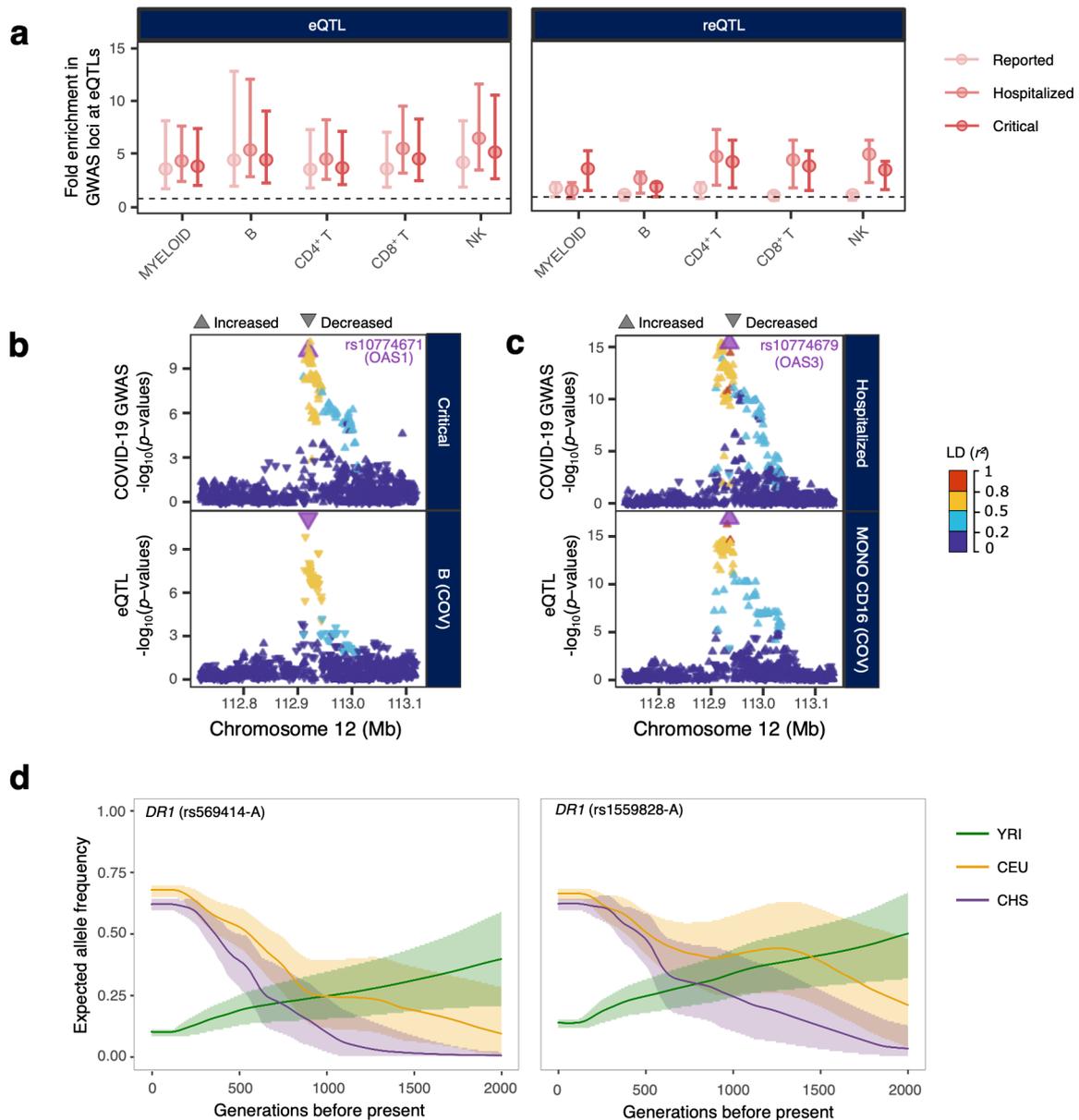
1798 according to the date on which selection began (gray for neutral). **e**, Fold enrichments in
1799 signals of positive selection (i.e., strong PBS) across the 22 cell types in Central Africans
1800 (YRI), West Europeans (CEU) and East Asians (CHS). Fold enrichments were calculated by
1801 genome-wide resampling of SNPs matched for minor allele frequency, LD and distance to the
1802 nearest gene. Vertical bars indicate 95% confidence intervals. **f**, Allele frequency trajectories
1803 over the past 2,000 generations in YRI (green) and CEU (yellow) of the *GBP7* reQTL
1804 (rs1142888-G). Shaded areas indicate 95% confidence intervals. **g** and **h**, Estimated period of
1805 selection over the past 2,000 generations, for 148 and 279 SARS-CoV-2 reQTLs with
1806 significant evidence of natural selection in Central Africans and West Europeans, respectively
1807 (max. $|Z\text{-score}| > 3$). In both panels, variants presenting strong signals of positive selection
1808 (i.e., top 5% for PBS) are shown in color. The transparent rectangle highlights the period
1809 between 770 and 970 generations ago (i.e., 21.5-27.2 thousand years ago) associated with
1810 polygenic adaptation targeting host coronavirus-interacting proteins. Variants are ordered
1811 along the x -axis in descending order of time to the onset of natural selection.
1812



1813
1814

1815 **Supplementary Figure 10 | Neanderthal introgression at loci regulating gene expression**
 1816 **in different cell types.** **a.** Adaptively introgressed eQTLs of host defense genes. From left to
 1817 right: (i) effects of the introgressed allele on gene expression across immune lineages and
 1818 stimulation conditions, (ii) clinical and functional annotations of associated genes, (iii)
 1819 present-day population frequencies of the introgressed alleles, (iv) percentile of archaic allele
 1820 frequency at the locus (CEU and CHS; dark shades: top 1%, light shades: top 5%, and (v)
 1821 effects of the target allele on COVID-19 risk (infection, hospitalization, and critical state).
 1822 Arrows indicate the increase/decrease in gene expression or disease risk with each copy of the
 1823 introgressed allele. Opacity increases with significance. In the leftmost panel, arrow colors
 1824 indicate the stimulation condition (gray: NS, red: COV, blue: IAV). For each eQTL, the
 1825 introgressed allele is defined as the allele segregating with the archaic haplotype in Eurasians.

1826 **b**, Effects on gene expression of two loci presenting strong evidence of adaptive introgression
1827 (*OAS1*, *TLRI*). For each locus, eQTL effect size and 95% confidence intervals are shown
1828 across the 22 cell types and the three stimulation conditions. **c** and **e**, Frequency and nature of
1829 archaic alleles at two introgressed loci (*TNFSF13B* and *TRAF3IP3*). Each dot represents an
1830 archaic allele and is colored according to its presence in Vindija Neanderthal (orange),
1831 Denisova (green) or both (yellow). The y-axis reflects their frequency in CEU (yellow) or
1832 CHS (purple) populations. The eQTL index SNP is shown in red. **d**, The Neanderthal-
1833 introgressed eQTL at *TRAF3IP3* is apparent only in IAV-infected monocytes, and is not
1834 detected in bystander cells (stimulated but not infected). **f**, Effects of the introgressed eQTL at
1835 *TNFSF13B* in MAIT cells (i.e., the cell type with the largest effect size). For **d** and **f**, middle
1836 line: median; box limits: upper and lower quartiles; whiskers: 1.5× interquartile range; points:
1837 outliers.
1838



1839
1840

1841 **Supplementary Figure 11| Colocalization of eQTLs and reQTLs with COVID-19-**
 1842 **associated loci. a**, Enrichment in COVID-19-associated loci at eQTLs and reQTLs in each
 1843 major lineage. For each set of eQTLs and each COVID-19 phenotype, fold enrichment and
 1844 resampling-based 95% confidence intervals are displayed. **b** and **c**, Colocalization of eQTLs
 1845 with COVID-19 GWAS hits at the *OAS1-3* locus. For each eQTL, the upper panel shows the
 1846 $\log_{10} p$ -value profile for association with COVID-19 phenotypes and the lower panel
 1847 represents the profile of $\log_{10} p$ -values for association with expression in a representative cell
 1848 type. Arrows indicate the direction of the effect at each SNP. The color code reflects LD (r^2)
 1849 with the consensus SNP, shown in purple, identified by colocalization analysis. **d**, Allele
 1850 frequency trajectories over the last 2,000 generations in the three populations of two *DR1*
 1851 eQTLs (rs569414 and rs1559828) that colocalize with COVID-19 severity loci. Shaded areas
 1852 indicate 95% confidence intervals.