

1 **A frequent ancestral *NFKB1* variant predicts risk of infection or allergy**

2 A. Y. Chong¹, N. Brenner², A. Jimenez-Kaufmann³, A. Cortes⁴, M. Hill⁵, T. J. Littlejohns⁶, J. J. Gilchrist^{1,7}, B. P.
3 Fairfax⁸, J. C. Knight^{1,9}, F. Hodel^{10,11}, J. Fellay^{10,11,12}, G. McVean⁴, A. Moreno-Estrada³, T. Waterboer², A. V. S.
4 Hill^{1,13}, A. J. Mentzer^{1, 4}

5 ¹The Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

6 ²Division of Infections and Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg,
7 Germany

8 ³Advanced Genomics Unit, National Laboratory of Genomics for Biodiversity (LANGEBIO), CINVESTAV,
9 Irapuato, Mexico

10 ⁴Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford,
11 UK

12 ⁵MRC-Population Health Research Unit, University of Oxford, Oxford, UK

13 ⁶Nuffield Department of Population Health, University of Oxford, Oxford, UK

14 ⁷Department of Paediatrics, University of Oxford, Oxford, UK

15 ⁸Department of Oncology, University of Oxford, Oxford, UK

16 ⁹NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

17 ¹⁰Global Health Institute, School of Life Sciences, EPFL, Lausanne, Switzerland

18 ¹¹Swiss Institute of Bioinformatics, Lausanne, Switzerland

19 ¹²Precision Medicine Unit, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

20 ¹³The Jenner Institute, University of Oxford, Oxford, UK

21

22 **Abstract**

23 Infectious agents contribute significantly to the global burden of diseases, through both acute infection and
24 their chronic sequelae. We leveraged the UK Biobank to identify genetic loci that influence humoral
25 immune response to multiple infections. From 45 genome-wide association studies in 9,611 participants
26 from UK Biobank, we identified *NFKB1* as a locus associated with quantitative antibody responses to
27 multiple pathogens including those from the herpes, retro- and polyoma-virus families. An insertion-
28 deletion variant thought to affect *NFKB1* expression (rs28362491), was mapped as the likely causal variant.
29 This variant has persisted throughout hominid evolution and could play a key role in regulation of the
30 immune response. Using 121 infection and inflammation related traits in 487,297 UK Biobank participants,
31 we show that the deletion allele was associated with an increased risk of infection from diverse pathogens
32 but had a protective effect against allergic disease. We propose that altered expression of *NFKB1*, as a

33 **NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
result of the deletion, modulates haematopoietic pathways, and likely impacts cell survival, antibody

34 production, and inflammation. Taken together, we show that disruptions to the tightly regulated immune
35 processes may tip the balance between exacerbated immune responses and allergy, or increased risk of
36 infection and impaired resolution of inflammation.

37

38 **Main text**

39 There is significant evidence that human genetic variation influences susceptibility to acute disease caused
40 by infectious pathogens. Case-control genome-wide association studies (GWAS) of a variety of infections
41 such as human immunodeficiency virus, hepatitis B and C viruses, leprosy, typhoidal and non-typhoidal
42 *Salmonella*, meningococcal disease, and more recently SARS-CoV-2 have identified a range of variants
43 spanning pathogen recognition, immune activation and immunologic memory robustly associated with
44 differential susceptibility¹⁻¹⁰. Furthermore, using large biobanks of human genetic data linked with
45 participant self-report, studies such as 23andMe, are replicating and building on these findings from case-
46 control analyses of infection susceptibility¹¹. Together, these studies are helping improve our
47 understanding of how these altered responses to infection may influence our susceptibility to future
48 sequelae of infection such as meningitis, mononucleosis, plantar warts and rheumatic heart disease.
49 Despite these advances, there remain challenges to using both study approaches. Firstly, achieving
50 sufficient power to identify novel genetic signals in case-control studies is limited by challenges in recruiting
51 large numbers of cases, and both study approaches may be affected by pathogen heterogeneity that may
52 impact on signatures of genetic association, whilst the self-report approach will be influenced by limitations
53 in accuracy and recall of self-reporting.

54 Antibodies are a critical component of response to infection and represent a marker of prior exposure or
55 chronic carriage with agents capable of latent or chronic infection states. Now that methods are available
56 to accurately measure multiple antibodies simultaneously in large numbers of individuals, these represent
57 stable biological markers that can be used to understand disease risk. Furthermore, there is evidence from
58 vaccine studies that the magnitude of antibody response and levels of circulating antibody may reflect the
59 likelihood of protection against either primary or subsequent disease, although this may depend on the
60 antigen to which the antibody is directed, and the functional impact of antibody binding and activity¹².

61 Human genetic variation is recognised to significantly influence total antibody levels¹³. Genetic analyses of
62 antibody response levels against vaccine preventable infections, such as hepatitis B virus, have
63 demonstrated a strong influence of genetic loci, including the major histocompatibility locus (MHC) in
64 predicting antibody magnitude overlapping with GWAS of disease susceptibility¹⁴. The MHC has also been
65 implicated in varied antibody responses to infections including JC virus, and influenza¹⁵⁻¹⁷. Furthermore,

66 primary immunodeficiencies such as common variable immunodeficiency (CVID) and X-linked
67 agammaglobulinaemia are well recognised disorders that increase the risk of multiple infections owing to
68 impaired humoral function and antibody deficiencies^{18–20}.

69 The availability of antibody data against 45 antigens representing 20 infectious agents in combination with
70 genome-wide genotyping data in a subset of UK Biobank participants now offers the opportunity to
71 investigate whether there are common variants that influence both response to antibodies and subsequent
72 susceptibility to both acute infectious disease and their chronic sequelae. While previous studies have
73 identified genetic associations with single antigens or pathogens using this UK Biobank dataset^{17,21}, we
74 hypothesise that there are genetic variants and loci that have a common effect on antibody responses to
75 multiple infections through effects on central immunological pathways. To identify genetic loci that
76 influence antibody titres against multiple infections, we performed GWAS for magnitude of antibody
77 responses to each antigen and identified *NFKB1* as a novel and distinct locus of interest, on contrast to the
78 other, more familiar, loci that were identified including immunoglobulin genes, the major
79 histocompatibility complex and *FUT2*. Given the role of *NFKB1* in inflammation and the induction and
80 regulation of immune responses to infection, we further investigated the association between *NFKB1*
81 variation and infectious and inflammatory diseases using data from 487,297 individuals. Finally, to better
82 understand the impact of variation in this locus on underlying immune mechanisms, we explored the effect
83 of this variant on blood cell traits and cellular expression data both within UK Biobank and in dedicated
84 expression quantitative trait datasets.

85

86 *NFKB1* is associated with antibody responses against multiple pathogens

87 We have previously described the use of a Multiplex Serology panel to measure immunoglobulin G (IgG)
88 antibody responses against 45 antigens from 20 infectious agents implicated in the pathogenesis of chronic
89 disease, applied to 9,695 individuals in UK Biobank⁽²²⁾; Supplementary Tables 1 and 2 of this manuscript).
90 Although we observed that antibody responses to antigens from the same pathogen were more strongly
91 correlated with each other than with those of other pathogens (Supplementary Fig. 1), we did observe all
92 antibody responses to be weakly correlated with each other (Spearman's rho -0.05- 0.86). It is likely that
93 shared genetic variants and environmental factors influence these correlated responses. Using the 9,611
94 individuals with both serology data and imputed genetic data (Supplementary Tables 1 and 2), we
95 performed genome-wide association studies of quantitative antibody responses to 45 antigens to identify
96 these common genetic loci that influence response to multiple infections.

97 We identified 27 genome-wide significant ($P < 5 \times 10^{-8}$) loci associated with magnitude of antibody
98 responses to any antigen (Fig. 1; Extended data Table 1). We observed four genomic regions, localising to
99 *NFKB1* (most significant association with HHV-6 (IE1A); beta = -0.10; $P = 1.30 \times 10^{-10}$), the extended MHC
100 region (EBV (EBNA1); beta = -0.32; $P = 6.90 \times 10^{-97}$) the IGH locus (HPV16 (E6); beta = 0.19; $P = 1.60 \times 10^{-10}$),
101 and *FUT2* (JCV (VP1); beta = -0.13; $P = 1.70 \times 10^{-21}$), which were associated with antibody response to
102 multiple antigens, and where at least one association signal passed Bonferroni correction ($P < 1.1 \times 10^{-9}$).
103 While all four genetic regions are well recognised to drive differences in infection susceptibility, this is the
104 first GWAS to identify *NFKB1* in relation to antibody responses against infectious disease ^{6,11,14–17,21,23}.

105 Associations were observed within the extended MHC region for at least one antigen from each of the 20
106 pathogens, with associations for human cytomegalovirus (CMV), *Chlamydia trachomatis*, Epstein-Barr virus
107 (EBV), human herpesvirus 6 and 7 (HHV-6 and 7), human papillomavirus type 16 (HPV16), herpes simplex
108 virus types 1 and 2 (HSV-1 and -2), human T-lymphotropic virus 1 (HTLV-1), JC polyomavirus (JCV), Merkel
109 cell polyomavirus (MCV), *Toxoplasma gondii* (*T. gondii*), and varicella-zoster virus (VZV) reaching genome-
110 wide significance (Extended Data Table 2). This suggests that, with adequate power, we are now able to
111 resolve HLA associations with responses to antigens considered most clinically relevant (such as EBNA-1
112 and EA for EBV; Fig. 1a inset), such as those described from the same UK Biobank cohort ^{16,17,21}.

113 Of particular interest is the identification of multiple variants in *NFKB1* as a common locus associated with
114 antibody responses across multiple pathogens and pathogen types, including EBV, human papillomavirus
115 type 18 (HPV18), HIV, HTLV-1, Kaposi's sarcoma-associated herpesvirus (KSHV), and *T. gondii*. To confirm
116 the likely contribution of this locus to multiple infectious antigen responses we replicated this association
117 demonstrating a consistent direction of effect for three antigens (EBV EAD, EBV ZEBRA and HHV-7 U14)
118 using an independently recruited cohort (CoLaus/PsyCoLaus; see Supplementary Table 2, Extended Data
119 Table 4, ¹⁶) where data using the same Multiplex Serology technology was available but with a different set
120 of antigens (Fig. 1b). *NFKB1* is one of five genes that encode the NF- κ B family of transcription factors. The
121 NF- κ B family plays a critical role in induction and mediation of pro-inflammatory response to infection, and
122 for development and maintenance of blood cells, and immune tissues through modulation of apoptosis.
123 Transcriptomic profiling of immune response to pathogens suggests that *NFKB1* forms part of a larger set of
124 genes involved in regulation of immune response to a wide range of pathogen infections ²⁴. Furthermore,
125 mutations in *NFKB1* have been identified as one of the most common causes of monogenic CVID in
126 European populations ^{18,20}.

127
128 Ancestral variation in *NFKB1* is under balancing selection in humans and may have conferred evolutionary
129 benefits in early hominids

130 Using a combination of fine mapping and examination of possible variant effects, we identified the most
131 likely causative variant to be rs28362491, a 4bp insertion-deletion (indel) variant within a promoter region
132 upstream of *NFKB1* which has been associated with multiple disease phenotypes in a range of previously
133 published genetic association studies. Examples of such traits with significant evidence of effect include
134 coronary artery disease with consistent reports of increased odds of disease with presence of the deletion
135 demonstrated in European (Odds ratio (OR) = 2.88 (95% CI 1.21-6.84), Extended Data Table 3), Indian (OR =
136 1.26 (1.03-1.55)) and Uyghur (OR = 1.58 (1.22-2.05)) populations; risk of lung injury in acute respiratory
137 distress syndrome in Europeans (OR = 3.7 (1.8-7.9)); and ulcerative colitis in Europeans (OR = 1.57 (1.14-
138 2.16)). Furthermore, the same region has been found to be associated in the same direction for other traits
139 where although the rs28362491 variant was not reported, it is likely that the effect may co-localise. Traits
140 of interest include tonsillectomy (OR = 1.08 (1.06-1.08)), ankylosing spondylitis (OR = 1.12 (1.08-1.16)),
141 primary biliary cirrhosis (OR = 1.34 (1.11-1.23)) and mouth ulcers (OR = 1.03 (1.02-1.04)) amongst others
142 (Extended Data Table 3). Intriguingly the same locus and/or variant was associated with other traits in the
143 opposite direction including allergic rhinitis (OR = 0.96 (0.95-0.97)), hayfever and eczema (OR = 0.96 (0.95-
144 0.97)) and cancer risk (OR = 0.89 (0.83-0.96)). The rs28362491 deletion is proposed to disrupt a promoter
145 binding site, lowering *NFKB1* expression²⁵. Given the broad range of effects of this allele and locus on
146 diverse traits in worldwide populations we explored global population frequencies of the deletion allele
147 (rs28362491:delATTG) using data from the 1000 Genomes Project and found evidence using Tajima's D test
148 for non-neutrality^{26,27} that variation within this locus is maintained through balancing selection, (Fig. 2b,c).
149 Notably, Tajima's D was negative for West Africans (YRI, D = -1.34 - -0.81), where allele frequency of the
150 deletion allele was higher than global average, than for other populations (CEU: D = 0.37-1.58, CHB: D =
151 0.67 - 2.05) suggesting that this locus is still undergoing purifying selection in some population groups.

152 To identify the ancestral source of the variant, we compared the region immediately upstream of *NFKB1*
153 across primates and early hominids and determined that the deletion allele is the ancestral variant (Fig. 2d),
154 with the insertion arising early in hominid evolution between 550 ka and 8.4 million years ago. This was
155 some time after the divergence of early hominids from chimpanzees (4.7–8.4 million years; ^{28,29}), but before
156 the divergence of archaic and modern humans (765–550 ka; ³⁰). That this variant has been maintained
157 throughout hominid evolution suggests that the insertion, and the resulting promoter binding region are
158 evolutionarily significant, further supporting the presence of balancing selection acting on this locus.

159

160 *NFKB1* is associated with dysregulation of inflammatory responses, balancing susceptibility to infections
161 with risk of allergy

162 We found that the direction of effect for rs28362491 was consistent across 9 antigens for which we
163 observed a signal for serum antibody responses in UK Biobank (EBV (EAD and ZEBRA), HHV-6 (IE1A), HHV-7
164 (U14), HPV-18 (L1), HTLV-1 (env), KSHV (K8.1), *T. gondii* (sag1)), with the deletion allele
165 (rs28362491:delATTG) being associated with lower antibody responses in UK Biobank. Given this
166 observation, with our observed replication in CoLaus/PsyCoLaus, and the published associations with
167 multiple infectious or inflammatory conditions (Extended Data Table 3), we further tested for the effect of
168 rs28362491:delATTG specifically on susceptibility to infection or inflammatory disease in UK Biobank, using
169 prevalent and incident hospital inpatient (ICD-10) or self-reported cases of infection and inflammatory
170 conditions in 487,297 participants (case definitions outlined in Supplementary Tables 3 and 4). Firstly, we
171 replicated the previously published associations with diverse traits such as allergic rhinitis and Ulcerative
172 colitis as reported in Extended Data Table 3 confirming the variant's role in a variety of traits (Fig. 3A). Then
173 we explored the effect of the variant on reported traits related to the measured antibody responses. For
174 example we found that although rs28362491:delATTG was consistently associated with reduced antibody
175 levels against all measured HPV antigens, there was an increased odds of reporting cervical intraepithelial
176 neoplasia (OR = 1.17 (1.02-1.18)) or having an ICD code of papillomavirus (OR = 1.09 (1.02-1.34)) (Fig. 3B).
177 We observed a similar pattern for EBV where the deletion was associated with reduced responses against
178 multiple EBV antigens (e.g. EA-D: OR = 0.94 (0.92-0.97)), yet there was a trend towards increased odds of
179 self-report of EBV infection/infectious mononucleosis (OR = 1.07 (0.98-1.18)) and multiple sclerosis (which
180 has been plausibly causally linked with EBV infection; OR = 1.05 (0.98-1.11) in our UKB data), as well as a
181 similar effect for *H. pylori*.

182 We then proceeded to test the association of the variant with a wider number of infectious and
183 inflammatory diagnoses. Out of 39 infections tested, the deletion allele increased risk of infection in three,
184 and decreased the risk of infection in one (*Neisseria meningitidis* (ICD), OR = 0.704 (0.509-0.974); $P =$
185 3.43×10^{-2}). Since many infections were limited in number, we also tested the effect of the deletion allele
186 against a phenotype consisting of any infection and found that overall there was a significant increase in
187 risk associated with the deletion (OR = 1.014 (1.001-1.027), $P = 3.31 \times 10^{-2}$; Supplementary Table 5).
188 Furthermore, we found that the same variant was associated with an increased risk of diseases associated
189 with dysregulation of inflammatory responses and chronic inflammation, such as psoriasis. In contrast, we
190 found that the deletion allele was protective against allergic disease such as allergic rhinitis and asthma
191 which are associated with acute inflammatory responses.

192 Altogether, these results demonstrate that this variant is associated with diverse groups of infectious or
193 allergic diagnoses. It appears that the deletion tends to increase the risk of infectious or chronic
194 inflammatory diagnoses, possibly associated with reduced antibody responses. Conversely, the insertion is
195 associated with increased antibody response (and thus reduced risk of infectious diagnoses) but will come

196 at an increased risk of allergic conditions. Given these observed divergent patterns of association with risk
197 of inflammatory and infectious or allergic disease, alongside of our observation of balancing selection, we
198 proceeded to test for risk of mortality in UK Biobank participants due to allergic, infectious or inflammatory
199 disease. Although low in frequency given the relative immaturity of the UK Biobank cohort study, we
200 observed similar trends of a protective effect of the deletion variant against death from allergic disease (OR
201 = 0.846 (0.658-1.088), $P = 0.192$) and an increased risk of infection (OR = 1.065 (0.982-1.155), $P = 0.129$)
202 and inflammation (OR = 1.038 (0.991-1.087), $P = 0.117$) (Supplementary Table 6).

203

204 *NFKB1* might act as a master regulator of cellular transcription in immune cells

205 Having found significant evidence of the rs28362491:delATTG variant being associated with phenotypic
206 disease endpoints and our intermediate antibody traits, we next sought to use these associations to
207 understand the likely molecular pathways disrupted leading to increased infection risk. We first
208 investigated the effect of rs28362491:delATTG on white blood cell counts and function (Supplementary
209 Table 7). Using cell count data from UK Biobank, we observed that this variant was not significantly
210 associated with lower overall white blood cell counts ($P = 0.96$), but instead had a higher total number of
211 lymphocytes (beta = 1.98×10^{-2} , $P = 3.30 \times 10^{-24}$) and eosinophils (beta = 6.49×10^{-3} , $P = 9.30 \times 10^{-4}$) than those
212 without. The deletion carriers also had lower counts of basophils (beta = -1.25×10^{-2} , $P = 2.00 \times 10^{-9}$),
213 monocytes (beta = -1.50×10^{-2} , $P = 2.70 \times 10^{-15}$), neutrophils (beta = -7.27×10^{-3} , $P = 2.50 \times 10^{-4}$), and platelets
214 (beta = -5.79×10^{-3} , $P = 1.30 \times 10^{-3}$) (Fig. 4).

215 We next interrogated the differential expression of *NFKB1* in diverse blood cell types and found that the
216 effect varied depending on the stimulation state of the cell. As rs28362491 is upstream of *NFKB1*, we used
217 a tightly linked SNP (rs72696119, $r^2 = 0.999$ $D' = 0.999$) in the 5' untranslated region (UTR) of *NFKB1* as a
218 tagging SNP for eQTL analyses. We found that rs72696119 is a cis-eQTL for *NFKB1* in resting monocytes ($P =$
219 1.20×10^{-2}), neutrophils ($P = 8.00 \times 10^{-3}$), NK cells ($P = 9.23 \times 10^{-4}$) and naive CD4+ and CD8+ T-cells ($P =$
220 3.50×10^{-4} and $P = 1.40 \times 10^{-3}$) but not in B-cells ($P = 0.99$). All significant loci in all tested analyses were found
221 to colocalise.

222

223 Summary

224 The genetic architecture underlying susceptibility to a host of infectious diseases is being resolved to
225 greater detail. Loci including the MHC, immunoglobulin loci and FUT2 secretor status are becoming
226 associated with a larger number of infectious traits when detected through multiple study formats, either

227 through case-control analyses of clinical disease definition, self-report or antigen-specific diagnostics. Here
228 we identify rs28362491 in the promoter region of *NFKB1*, which is common in all populations (global allele
229 frequency (del) = 0.42), that incurs a subtle, yet demonstrable effect across both the UK Biobank population
230 and, where available, other population datasets, modulating the risk of a range of infections. We found that
231 this ancestral *NFKB1* variant predates speciation of early humans and has been maintained in human
232 populations at a high frequency since with significant evidence of balancing selection occurring over long
233 evolutionary time. As expected, given the known pleiotropy of NF- κ B signalling, our study provides
234 significant evidence of the delicate balance between immune response to infection and inflammatory
235 processes, where disruptions can tip the balance towards exacerbated immune responses and allergy or
236 towards increased risk of infection and impaired resolution of inflammation. There are likely multiple
237 mechanisms underlying this effect, involving differential haematopoiesis and immunoglobulin class
238 switching driven by modulation of *NFKB1* gene expression.

239 On the basis of our functional work, we suggest that reduction of *NFKB1* expression, driven by the deletion
240 variant, favours development of lymphoid progenitor cells and lymphoid cell lineages over myeloid lineages
241 during haematopoiesis. NF- κ B activity regulates B-cell development and maturation and is important for T-
242 cell co-stimulation³¹. Despite moderate increases in the relative number of circulating lymphocytes,
243 reduced *NFKB1* expression in T-cells may result in impaired T-cell function, thereby making them less
244 responsive to infection. Variants tightly linked to rs28362491 have been associated with a reduced
245 IgA/IgG:IgM ratio (Extended data Table 3), suggesting that reduced *NFKB1* expression may also affect B-cell
246 class switching during activation^{13,20}, leading to the reduced IgG antibody levels observed in our study.

247 Interestingly, carriage of rs28362491:delATTG, was associated with higher levels of *NFKB1* gene expression
248 in all cell types except B and T cells (Fig. 4 insets). This was unexpected given that the deletion allele results
249 in the disruption of a promoter binding site. This effect was reversed in monocytes stimulated with
250 interferon gamma (IFN γ) or lipopolysaccharide (LPS), where lower *NFKB1* expression was observed after
251 stimulation. Although the mechanisms for this change in expression profile are unclear, it might also be
252 related to *NFKB1* self-regulation, as with the dysregulation of inflammation, where disruption of *NFKB1*
253 transcription disrupts NF- κ B activity, resulting in altered expression profiles in individuals carrying the
254 deletion allele. Our results support the role of *NFKB1* as a master regulator of cellular transcription that is
255 finely balanced depending on cell state and stimulation.

256 The observed inverse effect in allergy compared with infection and inflammatory disease may be a result of
257 impaired resolution of inflammation stemming from a dysregulation of *NFKB1* expression as a result of
258 disrupted promoter binding. NF- κ B family members and downstream gene products play a key role in
259 induction and regulation of inflammatory responses, and as such, are tightly regulated, possibly by relative
260 ratios of the NF- κ B p50-p65 (*NFKB1* encodes p50, p65 is encoded by *RELA*) and p50-p50 dimers³². Taken

261 together with our cellular findings, it would appear that in the absence of haploinsufficiency, variation and
262 regulation of expression is more important than overall *NFKB1* activity in altering risk of infection or
263 inflammatory disease. However, follow-up studies will be required to fully dissect the effects of mild
264 disruptions to NF- κ B pathways particularly since many of our observations are based on cross-sectional
265 analyses and therefore longitudinal analyses should be prioritised for the future.

266 Although we present evidence from a variety of sources and study formats in support of our observation,
267 that strengthens the credibility of the association, our work does not prove definitively that, firstly,
268 rs28362491:delATTG is the primary driving variant, nor that it is causal in all phenotypes studied. However,
269 our fine mapping methods would make it the most likely candidate and historic transcription binding
270 studies would make it the most plausible functional modulator of expression. Moreover, the subtle yet
271 extensive pleiotropy observed with carriage of the alternate alleles of this variant makes contemporary
272 methods of causal inference such as Mendelian Randomisation unreliable, and deeper validation of our
273 hypothesis of mechanism highly challenging to dissect.

274 Identifying rs28362491:delATTG as a robust association with the multiple traits presented here has
275 required data from a substantial number of individuals demonstrating that, although the effect of the
276 variant at an individual scale may be subtle, the variation is very important on a population level influencing
277 susceptibility to a range of diseases important to human health. These findings are likely to be important
278 for future disease prediction scores and potentially from a therapeutics perspective building on our
279 knowledge of the NFKB gene complex from primary immunodeficiency work where rare but large effect
280 variants contribute to disease susceptibility. However, the NF- κ B pathway will be difficult to target
281 therapeutically since an altered balance will encourage either the development of infection or allergy.
282 However, an acute disruption in favour of increased expression in cases of, for example adjuvantation
283 during vaccination, could offer an opportunity for some infectious disease prevention on a public health
284 scale.

285

286 **Methods**

287 Serology and GWAS

288 We used a Multiplex Serology platform described in Mentzer et. al.²² to measure antibody responses
289 against 45 antigens from 20 viral, bacterial, and protozoan pathogens present in the UK³³(Supplementary
290 Table 1). Viral pathogens included on the panel included herpes simplex virus types 1 and 2 (HSV-1 and 2),
291 varicella-zoster virus (VZV), Epstein-Barr virus (EBV), human cytomegalovirus (CMV), human herpesvirus 6
292 and 7 (HHV-6 and 7), Kaposi's sarcoma-associated herpesvirus (KSHV), hepatitis B virus (HBV), hepatitis C

293 virus (HCV), human T-lymphotropic virus 1 (HTLV-1), human immunodeficiency virus 1 (HIV-1), BK and JC
294 polyomaviruses (BKV and JCV), Merkel cell polyomavirus (MCV), and human papillomavirus 16 and 18
295 (HPV16 and 18). Other pathogens assayed were the bacterial pathogens *Chlamydia trachomatis* and
296 *Helicobacter pylori*, and protozoan parasite *Toxoplasma gondii*.

297 For all association tests, we used the genotyped and imputed autosomal variant datasets available for the
298 UK Biobank cohort ³⁴. This consists of 805,426 high-quality genotyped variants, generated from the UK
299 BiLEVE and UK Biobank Axiom arrays, and 93,095,623 imputed autosomal variants imputed using the
300 Haplotype Reference Consortium (HRC), UK10K and 1000 Genomes phase 3 reference panels ^{35–38}. For the
301 serology phenotypes, we selected a subset of 9,611 random, unrelated individuals that had genetic data
302 and serum samples available. Antibody responses across all participants regardless of serostatus were
303 normalised using a rank-based inverse normal transformation. We measured the Spearman’s rank
304 correlation between traits to determine whether there was any evidence of correlation between antibody
305 traits (Supplementary Fig. 1).

306 Association analyses using imputed genotype data and both the untransformed and normalised data from
307 quantitative serology traits were carried out using linear mixed models as implemented in BOLT-LMM
308 (minor allele frequency (MAF) > 0.01, imputation info score > 0.3; ³⁹) including age at recruitment and
309 genetic sex as covariates. To identify the genomic regions and candidate genes associated with antibody
310 response to each of the tested antigens, we first identified SNPs associated with tested antigens based on
311 genome-wide significance ($P < 5 \times 10^{-8}$) and suggestive significance ($P < 1 \times 10^{-5}$). Starting with the most
312 significant SNPs (peak SNP), we defined associated regions for each antigen based on calculated linkage
313 disequilibrium from best-guess genotypes ($R^2 \geq 0.2$, calculated with PLINK; ⁴⁰) between the peak SNP and
314 other SNPs reaching suggestive significance for the dataset. This was repeated for each SNP that was not
315 linked with a previous peak SNP. We then merged overlapping regions across all traits to define
316 independent regions of association for further analysis.

317

318 Finemapping of GWAS signals

319 For regions containing SNPs with an association p-value less than the Bonferroni corrected p value ($P <$
320 1.1×10^{-9}), we then performed conditional analyses to determine if there were multiple independent signals
321 present within the region. Analyses were carried out in GCTA using the --mlma option and adjusting for
322 allelic dosage of the top SNP for each associated trait ⁴¹. These gene dosages were also used to test for
323 interdependence between these associated regions and other regions of interest.

324 FINEMAP was used (default settings, $R2 = 0.8$; ⁴²) to identify potential causal groups (Bayes factor ≥ 2) of
325 SNPs within each of these regions based on summary statistics for each of the associated traits. These were
326 then compared across antigens and merged to define haplotypes of SNPs based on shared SNPs and
327 direction of effect. Causal groups sharing SNPs were only merged if the direction of effect across all
328 antigens were in a consistent direction relative to the minor allele.

329 Potential causal SNPs were identified from haplotypes shared across multiple antigens. To do this, we first
330 checked the location of SNPs relative to the Ensembl GRCh37.p13 gene annotations to identify the nearest
331 genes. The predicted effects of each variant were then retrieved using the Ensembl Variant Effect Predictor
332 ⁴³.

333

334 Cohorte Lausannoise replication study and meta-analysis

335 The Cohorte Lausannoise (CoLaus/PsyCoLaus) study includes 6,188 individuals of European ancestry living
336 in Lausanne, Switzerland who were randomly selected from the general population. Recruitment and
337 sampling protocols are described by Firmann et. al. ⁴⁴. Genotyping data for 5,399 individuals were obtained
338 using the Affymetrix Axiom Biobank array. Following quality control steps to remove poorly performing
339 variants and samples, imputed genotypes were generated using the Sanger Imputation Service ³⁸.
340 Genotypes were phased using EAGLE2 (v2.0.5; ³⁷) and the HRC panel ³⁸, and imputation was carried out
341 using the HRC reference panel, 1000 Genomes Phase 3 data, and the UK10K reference panel ^{35,36}. Serology
342 phenotypes were generated using a similar Multiplex Serology Panel to the UK Biobank data as described
343 above ^{16,33}.

344 Genome-wide association analyses were carried out using 4216 individuals with imputed genotype data
345 and serology data, and linear mixed models as implemented in GCTA (1.91.3beta; ⁴¹) and thresholded high
346 quality imputed genotype data (imputation info score > 0.8 , $MAF > 0.01$, $P_{HWE} < 10 \times 10^{-7}$). Age, sex, and the
347 top three principal components were included as covariates. For antigens with a signal at *NFKB1* and that
348 were also available from the CoLaus/PsyCoLaus cohort, we performed GWAS that were common across the
349 two cohorts, we extracted the set of intersecting SNPs for meta-analysis using Metasoft (v2.0.1, ⁴⁵).

350

351 Infection and inflammation associations in UK Biobank

352 We used self-reported non-cancer illness (Field ID: 20002) and prevalent and incident hospital inpatient
353 ICD-10 codes (Field ID: 41270) to define cases and controls for further infection and inflammation related
354 phenotypes within UK Biobank. For self-reported conditions, cases and controls were drawn from

355 participants who reported the presence or absence of non-cancer related conditions, and had genetic data
356 available, excluding those who did not respond to the question. Controls for infection-related traits derived
357 from self-reported data were defined as individuals who had reported no major infections for infection-
358 related phenotypes with genetic data available. Controls for other immune-related traits were drawn from
359 all unaffected individuals with genetic data.

360 Similarly, potential cases and controls based on ICD-10 data were drawn from those individuals with
361 hospital inpatient diagnoses and genetic data available. Case definitions are outlined in Supplementary
362 Tables 3 and 4. Controls for ICD-10 derived phenotypes were drawn from individuals with hospital inpatient
363 and genetic data, who were not affected by the trait of interest. In the case of infection-related traits, the
364 selection was further restricted to those who had not reported a major infection, or with no ICD codes
365 relating to infection (e.g. Chapter I (A00-B99), Chapter X: Diseases of the respiratory system: Acute upper
366 respiratory infections (J00-J06), Influenza and pneumonia (J09-J18), Other acute lower respiratory
367 infections (J20-J22)).

368 Where equivalent phenotypes were present in both self-reported and ICD-10 data, a joint phenotype was
369 defined using all possible cases and controls. Similarly, summary phenotypes for allergy, infection, and
370 inflammation were defined using all available cases and controls. Due to the small numbers of cases across
371 many phenotypes, we used only those with more than 50 cases.

372 Association tests for these traits were carried out using linear mixed models as implemented in SAIGE (v
373 0.36.3.2; ⁴⁶), also including age at recruitment and genetic sex as covariates, except for traits where only
374 one sex was present in the cases, in which case sex was not included, but controls were drawn only from
375 individuals of that sex (see Supplementary Tables 3 and 4 for more information). Due to relatively small
376 numbers of individuals for which cause of death was available (approximately 25,000), we used logistic
377 regression to assess the relationship between rs28362491:delATTG and the contribution of infection or
378 inflammatory conditions to cause of death. Similar to the GWAS analysis described above, we used imputed
379 dosages of rs28362491:delATTG, with age and sex as covariates. We also restricted analysis to individuals in
380 the white British subset as defined by UK Biobank.

381

382 Association with blood cell counts

383 Association tests for blood cell counts were carried out on the full UK Biobank cohort using transformed cell
384 count data and imputed genotypes. Traits included red blood cell (RBC) count, white blood cell (WBC)
385 count, platelet, basophil, eosinophil, lymphocyte, monocyte, and neutrophil counts. Values were

386 normalised using a rank-based inverse normal transformation prior to analysis. Association tests were
387 carried out using BOLT-LMM including age at recruitment and genetic sex as covariates.

388

389 eQTL analysis

390 To assess the regulatory function of rs28362491, we used the tightly linked, 5'UTR SNP rs72696119. We
391 correlated rs72696119 genotype with *NFKB1* RNA expression in naïve and stimulated primary immune cell
392 subsets from healthy European adults, using data from published and unpublished expression quantitative
393 trait locus (eQTL) studies. CD56+CD3- NK cells, CD19+ B cells, CD14+ monocytes, CD16+ neutrophils were
394 separated, genotyped, and gene expression quantified as previously described⁴⁷⁻⁵⁰. Following quality
395 control and normalisation of gene expression data, we correlated rs72696119 genotype with *NFKB1* RNA
396 expression in each cell type: B-cells (n = 279), NK cells (n = 245), neutrophils (n = 101), naïve monocytes (n =
397 414), and stimulated monocytes (LPS 2 hours, n = 261; LPS 24 hours, n = 322; IFN γ 24 hours, n = 367).
398 *NFKB1* expression was correlated with genotype by linear regression and analysis of variance (ANOVA),
399 including the first 25 principal components of gene expression data in each cell-type/condition to account
400 for confounding variation. P-values are calculated with F-tests (1 d.f.). Statistical analysis was performed in
401 R. We used publicly available eQTL data from the DICE database for CD4+ and CD8+ T cells to examine
402 association of rs72696119 with *NFKB1* expression in naïve T-cells^(51; <https://dice-database.org/>).

403

404 Colocalisation of eQTL and GWAS signals

405 Colocalisation testing between eQTL signals and antibody and blood cell GWAS association signals was
406 carried out using coloc (v5.0.1,⁵²) with susieR (v0.11.92,^{53,54}) and the intersection of variants from all GWAS
407 phenotypes and the eQTL traits from the region immediately upstream of and across *NFKB1*. LD was
408 calculated independently for each dataset using PLINK. Colocalisation was tested between pairs of antibody
409 traits, between HHV6 (IE1A) and blood cell traits, between HHV6 (IE1A) and eQTL datasets, and between
410 related blood cell traits and eQTL datasets (e.g. monocyte count and unstimulated and stimulated
411 monocyte eQTL datasets) (Supplementary table 8).

412

413 *NFKB1* phylogenetics

414 We extracted the first exon of *NFKB1* and 500bp immediately upstream from 11 high quality primate
415 genomes present in Ensembl. These sequences were aligned using ClustalW⁵⁵ and manually inspected to

416 determine the presence of the 4bp indel variant. To confirm the absence of the insertion in great apes, we
417 used variation data for six great ape species: Eastern lowland gorilla (*Gorilla beringei*; n = 3), Western
418 lowland gorilla (*Gorilla gorilla*; n = 3), Bonobo (*Pan paniscus*, n = 13), Chimpanzee (*Pan troglodytes*; n = 14),
419 Sumatran orangutan (*Pongo abelii*; n = 5), and Bornean orangutan (*Pongo pygmaeus*; n = 5)⁵⁶. We also
420 used variation data from three Neanderthals, one Denisovan genome, and three ancient humans³⁰ to
421 estimate the latest common ancestor where the indel variant may have arisen.

422

423 Selection testing

424 Tajima's D population genetic summary statistics and rank scores for the CEU, YRI, and CHB populations
425 were retrieved from the Human Positive Selection Browser²⁶ for *NFKB1* and regions 200kb upstream and
426 downstream of the gene.

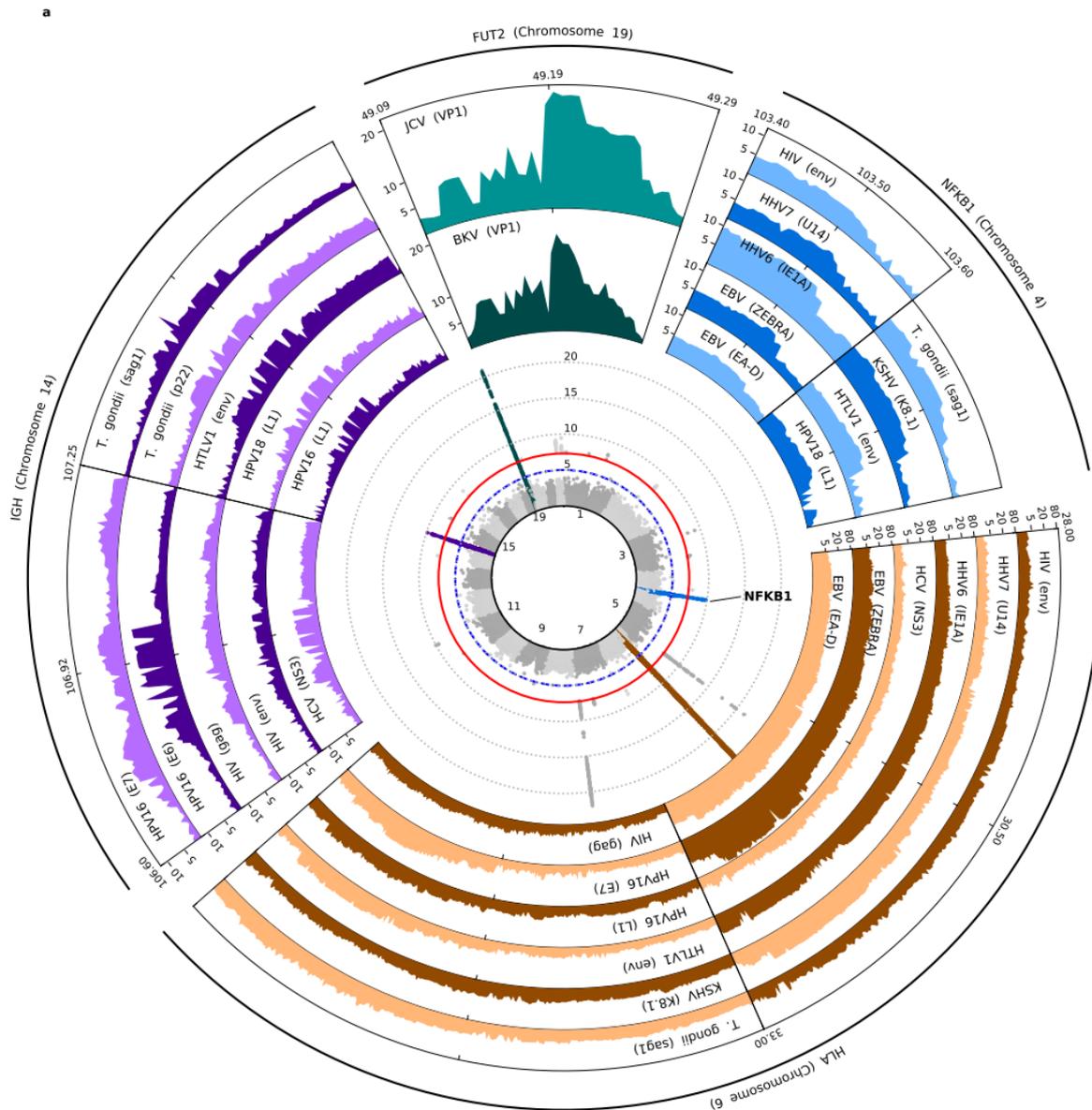
427

428

429

430

431 **Figures**



b

Phenotype	UK Biobank (n=9611)				CoLaus/PsyCoLaus (n=4216)				Meta-analysis		
	Seropositive	beta	SE	P	Seropositive	beta	SE	P	beta	SE	P
EBV (EAD)	8278	-0.059	0.014	3.20×10^{-5}	3257	-0.117	0.024	9.15×10^{-7}	-0.074	0.012	1.13×10^{-9}
EBV (ZEBRA)	8763	-0.062	0.014	1.30×10^{-5}	3718	-0.116	0.023	7.20×10^{-7}	-0.076	0.012	3.01×10^{-10}
HHV6 (IE1A)	7469	-0.086	0.015	2.90×10^{-9}	766	-0.030	0.024	2.03×10^{-1}	-0.071	0.012	1.03×10^{-8}
HHV7 (U14)	9103	-0.057	0.014	5.00×10^{-5}	2244	-0.055	0.023	1.81×10^{-2}	-0.057	0.012	2.68×10^{-6}
KSHV (K8.1)	305	-0.069	0.014	1.50×10^{-6}	58	0.035	0.024	1.47×10^{-1}	-0.042	0.012	7.40×10^{-4}
T. gondii (sag1)	2190	-0.063	0.014	1.20×10^{-5}	1328	-0.004	0.023	8.73×10^{-1}	-0.047	0.012	1.33×10^{-4}

432

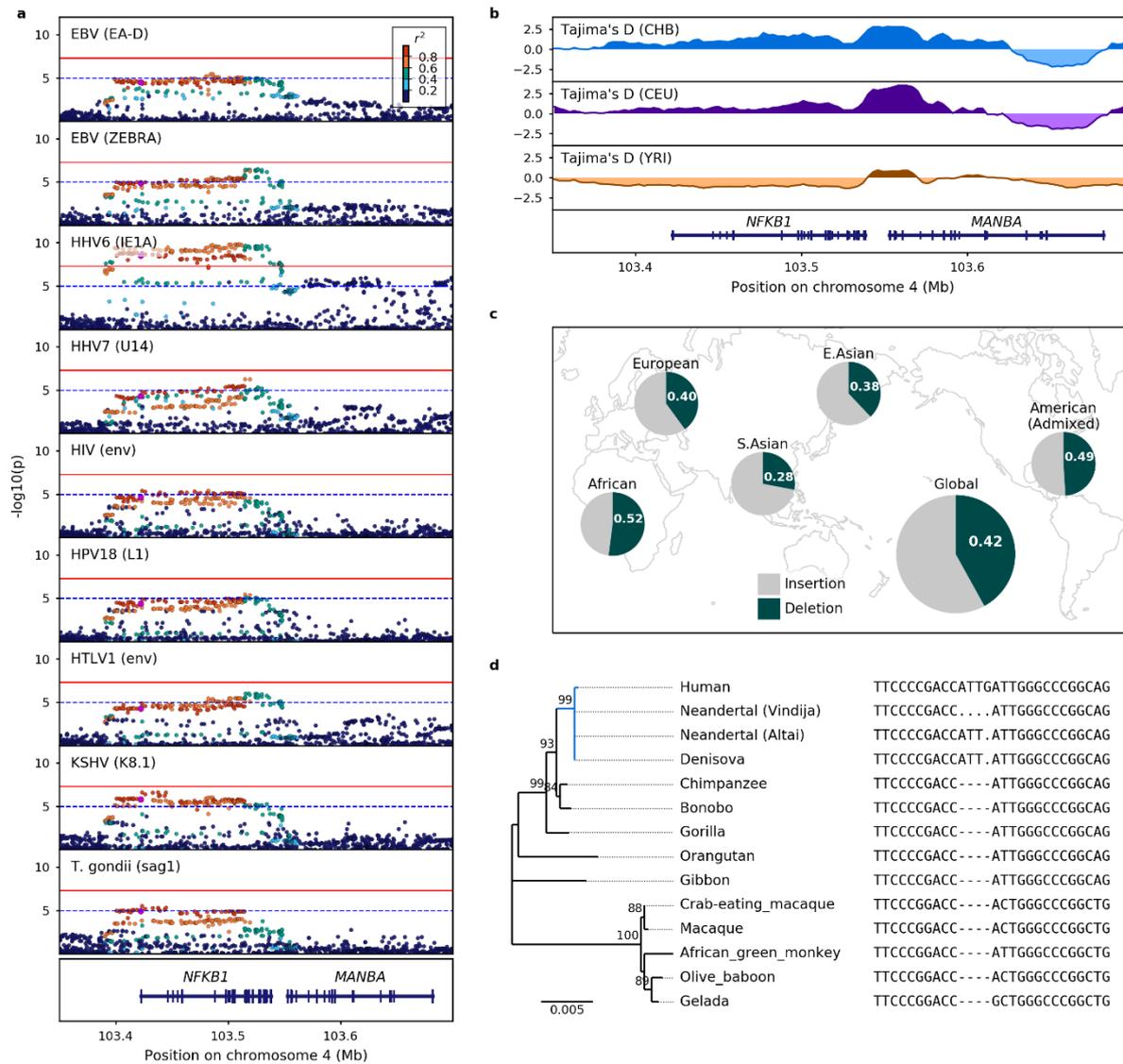
433 **Fig. 1: *NFKB1* is a key locus affecting antibody response, and risk of infection and inflammatory disease.**

434 (a) Central inset: stacked Manhattan plot of association statistics for antibody response to 45 pathogen-

435 derived antigens. Signals for the four loci with multiple associations are coloured: *NFKB1* (blue), the MHC

436 locus (brown), the IGH locus (purple), and *FUT2* (teal). Surrounding insets: Regional association plots for

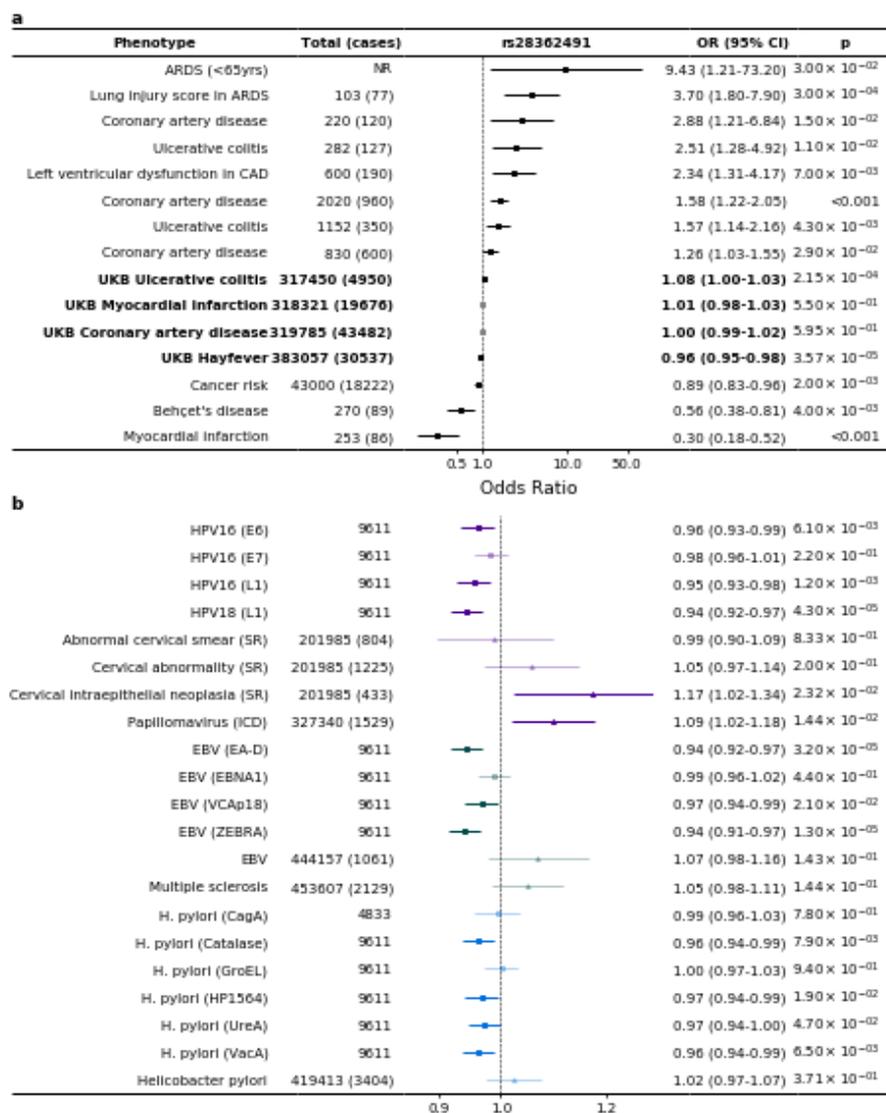
- 437 specific antigens demonstrating signals at each of the four highlighted loci. Y-axis (all plots) is $-\log_{10}(P)$ (b)
- 438 Meta-analysis of the NFKB1 locus for association with six antigens in common between the UK Biobank and
- 439 CoLaus/PsyCoLaus cohorts. Results for rs28362491 are presented.



440

441 **Fig. 2: rs28362491 is an ancient variant affecting antibody responses to a range of pathogens.** (a) Regional
442 association plots showing the consistency of associations across the *NFKB1*/*MANBA* region on chromosome
443 4 in UK Biobank. Variants are coloured based on linkage with rs28362491 (indicated in purple). The upper
444 red line indicates genome-wide significance ($P = 5 \times 10^{-8}$), and the lower blue dashed line indicates
445 suggestive significance ($P = 1 \times 10^{-5}$). (b) Binned Tajima's D selection statistics for 30 kb sliding windows
446 across *NFKB1*/*MANBA* for 1000 Genomes Han Chinese in Beijing (CHB), Central European (CEU), and Yoruba
447 in Ibadan, Nigeria (YRI) populations. A positive Tajima's D is indicative of balancing selection, while a
448 negative value suggests a recent selective sweep or population expansion. (c) Global population
449 frequencies of rs28362491 based on allele frequencies from the 1000 Genomes project. (d) Phylogeny of
450 *NFKB1* gene sequence across primates and hominids (highlighted in blue). The nucleotide sequence
451 surrounding the rs28362491 indel variant is shown alongside. Dashes indicate the ancestral deletion
452 variant; dots indicate potentially missing data from ancient DNA sequencing.

453

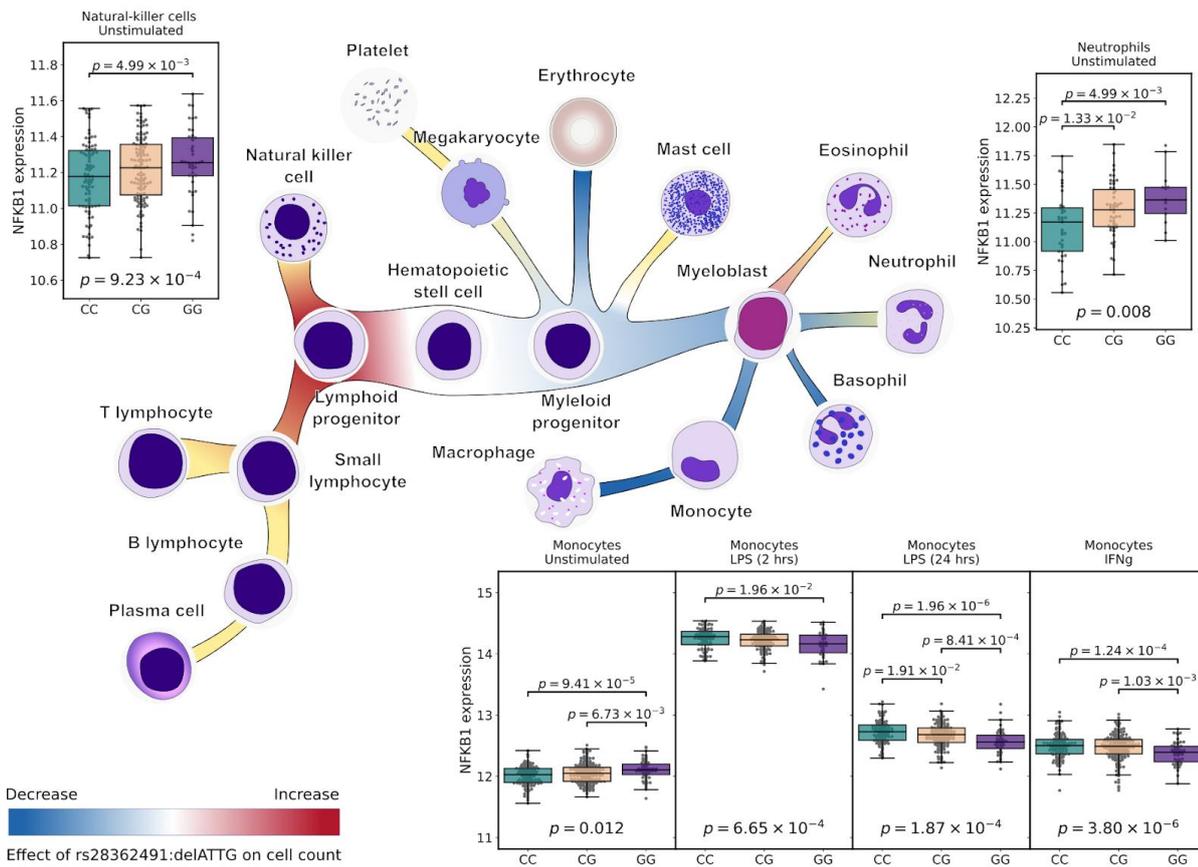


454

455 **Fig. 3: rs28362491:delATTG has a small, consistent effect**, associated with increasing the risk of a range of
 456 infections and disease associated with dysregulation of inflammation, while protecting against exacerbated
 457 immune responses and allergy. (a) Published associations between rs28362491 and disease, and
 458 comparable summary stats for UKB (highlighted in bold). References listed in Extended Data Table 3. (b)
 459 Odds ratios and 95% confidence intervals for association between rs28362491, normalised antibody
 460 response against 14 antigens and 7 health-record derived disease traits. Square markers indicate
 461 quantitative antibody response phenotypes, and triangle markers indicate phenotypes derived from health
 462 record data (ICD-10 and self-report). Solid colours indicate traits with a $P < 0.05$. ARDS: Acute respiratory
 463 distress syndrome; CAD: Coronary artery disease; ICD: Phenotype is derived from ICD-10 health record
 464 data; SR: Phenotype is derived from self-reported health record data.

465

466



467

468 **Fig. 4: rs28362491 variation is also associated with differential haematopoiesis and *NFKB1* expression in**
 469 **white blood cells.** Effect of rs28362491:delATTG on white blood cell counts is indicated by red-blue
 470 shading. Yellow shaded regions indicate a lack of information on cell counts (mast cells, B and T cells) or
 471 signals that did not reach genome-wide significance (NK cells, platelets, neutrophils, and eosinophils). Plot
 472 insets show differential expression by genotype at strongly linked 5`UTR SNP rs72696119 (delATTG linked
 473 with G), eQTL p-values, and p-values for pairwise T-tests for differential expression by genotype. Clockwise
 474 from top-left: Unstimulated NK-cells, unstimulated neutrophils, unstimulated and stimulated monocytes.
 475 Expression for B and T cells not shown.

476

477

478 References

- 479 1. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* 2021
480 1–8 (2021).
- 481 2. The Severe Covid-19 GWAS Group. Genomewide Association Study of Severe Covid-19 with
482 Respiratory Failure. *NEJM* **383**, 1522–1534 (2020).
- 483 3. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* 2020 591:7848
484 **591**, 92–98 (2020).
- 485 4. Davila, S. *et al.* Genome-wide association study identifies variants in the CFH region associated with
486 host susceptibility to meningococcal disease. *Nature Genetics* **42**, 772–776 (2010).
- 487 5. Duggal, P. *et al.* Genome-Wide Association Study of Spontaneous Resolution of Hepatitis C Virus
488 Infection: Data From Multiple Cohorts. *Annals of Internal Medicine* **158**, 235–245 (2013).
- 489 6. Dunstan, S. J. *et al.* Variation at HLA-DRB1 is associated with resistance to enteric fever. *Nature*
490 *Genetics* **46**, 1333–1336 (2014).
- 491 7. Gilchrist, J. J. *et al.* Risk of nontyphoidal Salmonella bacteraemia in African children is modified by
492 STAT4. *Nature Communications* **9**, (2018).
- 493 8. Vergara, C. *et al.* Multi-Ancestry Genome-Wide Association Study of Spontaneous Clearance of
494 Hepatitis C Virus. *Gastroenterology* **156**, 1496-1507.e7 (2019).
- 495 9. Wang, Z. *et al.* A large-scale genome-wide association and meta-analysis identified four novel
496 susceptibility loci for leprosy. *Nature Communications* **7**, 13760 (2016).
- 497 10. Zhang, F. *et al.* Identification of two new loci at IL23R and RAB32 that influence susceptibility to
498 leprosy. *Nature Genetics* **43**, 1247–1251 (2011).
- 499 11. Tian, C. *et al.* Genome-wide association and HLA region fine-mapping studies identify susceptibility loci
500 for multiple common infections. *Nature Communications* **8**, 599 (2017).
- 501 12. VanBlargan, L. A., Goo, L. & Pierson, T. C. Deconstructing the Antiviral Neutralizing-Antibody Response:
502 Implications for Vaccine Development and Immunity. *Microbiology and Molecular Biology Reviews* **80**,
503 989–1010 (2016).
- 504 13. Jonsson, S. *et al.* Identification of sequence variants influencing immunoglobulin levels. *Nat Genet* **49**,
505 1182–1191 (2017).
- 506 14. Png, E. *et al.* A genome-wide association study of hepatitis B vaccine response in an Indonesian
507 population reveals multiple independent risk variants in the HLA region. *Human Molecular Genetics*
508 **20**, 3893–3898 (2011).
- 509 15. Hammer, C. *et al.* Amino Acid Variation in HLA Class II Proteins Is a Major Determinant of Humoral
510 Response to Common Viruses. *Am J Hum Genet* **97**, 738–743 (2015).
- 511 16. Hodel, F. *et al.* Human genomics of the humoral immune response against polyomaviruses. *Virus*
512 *Evolution* **7**, 1–11 (2021).

- 513 17. Kachuri, L. *et al.* The landscape of host genetic factors involved in immune response to common viral
514 infections. *Genome Medicine* **12**, 1–18 (2020).
- 515 18. Hayden, M. S. & Ghosh, S. NF-kappaB in immunobiology. *Cell Res* **21**, 223–244 (2011).
- 516 19. Hayden, M. S., West, A. P. & Ghosh, S. NF-kappaB and the immune response. *Oncogene* **25**, 6758–6780
517 (2006).
- 518 20. Tuijnenburg, P. *et al.* Loss-of-function nuclear factor kappaB subunit 1 (NFKB1) variants are the most
519 common monogenic cause of common variable immunodeficiency in Europeans. *J Allergy Clin*
520 *Immunol* **142**, 1285–1296 (2018).
- 521 21. Butler-Laporte, G. *et al.* Genetic Determinants of Antibody-Mediated Immune Responses to Infectious
522 Diseases Agents: A Genome-Wide and HLA Association Study. *Open Forum Infect Dis* **7**, (2020).
- 523 22. Mentzer, A. J. *et al.* Identification of host–pathogen–disease relationships using a scalable multiplex
524 serology platform in UK Biobank. *Nature Communications* **13**, (2022).
- 525 23. Scepanovic, P. *et al.* Human genetic variants and age are the strongest predictors of humoral immune
526 responses to common pathogens and vaccines. *Genome Med* **10**, 59 (2018).
- 527 24. Jenner, R. G. & Young, R. A. Insights into host responses against pathogens from transcriptional
528 profiling. *Nat Rev Microbiol* **3**, 281–294 (2005).
- 529 25. Karban, A. S. *et al.* Functional annotation of a novel NFKB1 promoter polymorphism that increases risk
530 for ulcerative colitis. *Hum Mol Genet* **13**, 35–45 (2004).
- 531 26. Pybus, M. *et al.* 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of
532 natural selection in modern humans. *Nucleic Acids Res* **42**, D903-9 (2014).
- 533 27. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
534 *Genetics* **123**, 585–595 (1989).
- 535 28. Kumar, S., Filipski, A., Swarna, V., Walker, A. & Hedges, S. B. Placing confidence limits on the molecular
536 age of the human-chimpanzee divergence. *Proc Natl Acad Sci U S A* **102**, 18842–18847 (2005).
- 537 29. Patterson, N., Richter, D. J., Gnerre, S., Lander, E. S. & Reich, D. Genetic evidence for complex
538 speciation of humans and chimpanzees. *Nature* **441**, 1103–1108 (2006).
- 539 30. Prufer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science (1979)* **358**,
540 655–658 (2017).
- 541 31. Kaileh, M. & Sen, R. NF-kappaB function in B lymphocytes. *Immunol Rev* **246**, 254–271 (2012).
- 542 32. Mishra, A., Srivastava, A., Mittal, T., Garg, N. & Mittal, B. Role of inflammatory gene polymorphisms in
543 left ventricular dysfunction (LVD) susceptibility in coronary artery disease (CAD) patients. *Cytokine* **61**,
544 856–861 (2013).
- 545 33. Waterboer, T. *et al.* Multiplex Human Papillomavirus Serology Based on In Situ–Purified Glutathione S-
546 Transferase Fusion Proteins. *Clinical Chemistry* **51**, 1845–1853 (2005).

- 547 34. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–
548 209 (2018).
- 549 35. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**,
550 68–74 (2015).
- 551 36. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype
552 reference panel. *Nature Communications* **6**, 8111 (2015).
- 553 37. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nature*
554 *Genetics* **48**, 1443–1448 (2016).
- 555 38. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*
556 **48**, 1279–1283 (2016).
- 557 39. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts.
558 *Nature Genetics* **47**, 284–290 (2015).
- 559 40. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets.
560 *Gigascience* **4**, 7 (2015).
- 561 41. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait
562 analysis. *Am J Hum Genet* **88**, 76–82 (2011).
- 563 42. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide
564 association studies. *Bioinformatics* **32**, 1493–1501 (2016).
- 565 43. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 1–14 (2016).
- 566 44. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and
567 genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord* **8**,
568 6 (2008).
- 569 45. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of
570 genome-wide association studies. *Am J Hum Genet* **88**, 586–598 (2011).
- 571 46. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale
572 genetic association studies. *Nature Genetics* **50**, 1335–1341 (2018).
- 573 47. Naranbhai, V. *et al.* Genomic modulators of gene expression in human neutrophils. *Nature*
574 *Communications* **6**, 7545 (2015).
- 575 48. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific
576 master regulators and roles of HLA alleles. *Nature Genetics* **44**, 502–510 (2012).
- 577 49. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte
578 gene expression. *Science (1979)* **343**, 1246949 (2014).
- 579 50. Gilchrist, J. J. *et al.* Natural Killer cells demonstrate distinct eQTL and transcriptome-wide disease
580 associations, highlighting their role in autoimmunity. *bioRxiv* (2021) doi:10.1101/2021.05.10.443088.

- 581 51. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*
582 **175**, 1701-1715.e16 (2018).
- 583 52. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants.
584 *PLOS Genetics* **17**, e1009440 (2021).
- 585 53. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in
586 regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B*
587 (*Statistical Methodology*) **82**, 1273–1300 (2020).
- 588 54. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the “Sum of
589 Single Effects” model. *bioRxiv* (2021) doi:10.1101/2021.11.03.467167.
- 590 55. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive
591 multiple sequence alignment through sequence weighting, position-specific gap penalties and weight
592 matrix choice. *Nucleic Acids Res* **22**, 4673–4680 (1994).
- 593 56. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499**, 471–475
594 (2013).
- 595 57. Waage, J. *et al.* Genome-wide association and HLA fine-mapping studies identify risk loci and genetic
596 pathways underlying allergic rhinitis. *Nat Genet* **50**, 1072–1080 (2018).
- 597 58. Johansson, A., Rask-Andersen, M., Karlsson, T. & Ek, W. E. Genome-wide association analysis of 350
598 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema. *Hum Mol*
599 *Genet* **28**, 4022–4041 (2019).
- 600 59. Bajwa, E. K. *et al.* An NFKB1 promoter insertion/deletion polymorphism influences risk and outcome in
601 acute respiratory distress syndrome among Caucasians. *PLoS One* **6**, e19469 (2011).
- 602 60. Ellinghaus, D. *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and
603 highlights disease-specific patterns at shared loci. *Nat Genet* **48**, 510–518 (2016).
- 604 61. Yenmis, G. *et al.* Association of NFKB1 and NFKBIA polymorphisms in relation to susceptibility of
605 Behcet’s disease. *Scand J Immunol* **81**, 81–86 (2015).
- 606 62. Seidi, A., Mirzaahmadi, S., Mahmoodi, K. & Soleiman-Soltanpour, M. The association between NFKB1 -
607 94ATTG ins/del and NFKB1A 826C/T genetic variations and coronary artery disease risk. *Mol Biol Res*
608 *Commun* **7**, 17–24 (2018).
- 609 63. Lai, H.-M. *et al.* Genetic Variation in NFKB1 and NFKBIA and Susceptibility to Coronary Artery Disease in
610 a Chinese Uygur Population. *PLOS ONE* **10**, e0129144 (2015).
- 611 64. Adamzik, M. *et al.* Insertion/deletion polymorphism in the promoter of NFKB1 influences severity but
612 not mortality of acute respiratory distress syndrome. *Intensive Care Med* **33**, 1199–1203 (2007).
- 613 65. Dudding, T. *et al.* Genome wide analysis for mouth ulcers identifies associations at immune regulatory
614 loci. *Nat Commun* **10**, 1052 (2019).

- 615 66. Boccardi, V. *et al.* -94 ins/del ATTG NFKB1 gene variant is associated with lower susceptibility to
616 myocardial infarction. *Nutr Metab Cardiovasc Dis* **21**, 679–684 (2011).
- 617 67. Kawashima, M. *et al.* Genome-wide association studies identify PRKCB as a novel genetic susceptibility
618 locus for primary biliary cholangitis in the Japanese population. *Hum Mol Genet* **26**, 650–659 (2017).
- 619 68. Gonzalez-Serna, D. *et al.* Analysis of the genetic component of systemic sclerosis in Iranian and Turkish
620 populations through a genome-wide association study. *Rheumatology (Oxford)* **58**, 289–298 (2019).
- 621 69. Liou, Y. J. *et al.* Genome-wide association study of treatment refractory schizophrenia in Han Chinese.
622 *PLoS One* **7**, e33598 (2012).
- 623 70. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel
624 disease. *Nature* **491**, 119–124 (2012).
- 625 71. Borm, M. E., van Bodegraven, A. A., Mulder, C. J., Kraal, G. & Bouma, G. A NFKB1 promoter
626 polymorphism is involved in susceptibility to ulcerative colitis. *Int J Immunogenet* **32**, 401–405 (2005).
- 627 72. Wang, D. *et al.* Genetic association between NFKB1 -94 ins/del ATTG Promoter Polymorphism and
628 cancer risk: a meta-analysis of 42 case-control studies. *Sci Rep* **6**, 30220 (2016).
- 629 73. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to
630 complex human diseases. *Nature Genetics* **50**, 390–400 (2018).

631

632 **Data availability**

633 GWAS summary statistics are available on Zenodo. Serology data:

634 <https://doi.org/10.5281/zenodo.7347714>; Heath record and Blood cell count data:

635 <https://doi.org/10.5281/zenodo.7347792>

636

637 **Acknowledgements**

638 We thank the UK Biobank participants and all individuals involved in recruitment, data and sample
639 collection, data curation and release. This research has been conducted using the UK Biobank Resource
640 under Application Number 43920. This research has been conducted using data from the CoLaus/PsyCoLaus
641 Study.

642

643 AJM was supported by a Wellcome Trust Fellowship with reference 106289/Z/14/Z and the National
644 Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). AYC was supported by the
645 Research Councils UK Newton Fund Award with reference MR/N028937/1. AJ-K and AM-E were supported
646 by The Mexican Biobank Project (CONACYT-Newton Fund grant number FONCICYT/50/2016). JIG is funded

647 by a National Institute for Health Research (NIHR) Clinical Lectureship. JF was supported by the Swiss
648 National Science Foundation, grant #175603. The research was supported by the Wellcome Trust Core
649 Award Grant Number 203141/Z/16/Z with additional support from the NIHR Oxford BRC. The views
650 expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of
651 Health.

652
653

654 **Author contributions**

655 AYC and AJK carried out the GWAS of antibody traits; AYC performed additional analysis of *NFKB1*-related
656 traits. NB, MH, TJL, TW, and AJM performed the antibody data generation and NB, TJL, TW, and AJM
657 curated the data. AC, MH, TJL contributed to the methods development. JJG, BPF, and JCK carried out
658 analysis of the eQTL data. FH and JF carried out analysis of the CoLaus/PsyCoLaus cohort. TW, AVSH, and
659 AJM designed the experiments. GM, AME, TW, AVSH, and AJM supervised the research and provided
660 funding.

661

662 **Competing interests**

663 Nil

664

665

666

667