A common *NFKB1* variant detected through antibody analysis in UK Biobank predicts risk of infection and allergy

Authors

Amanda Y. Chong, Nicole Brenner,
Andres Jimenez-Kaufmann, ..., Tim Waterboer,
Adrian V.S. Hill, Alexander J. Mentzer

Correspondence

amanda.chong@well.ox.ac.uk (A.Y.C.),
alexander.mentzer@ndm.ox.ac.uk (A.J.M.)

Infectious agents contribute significantly to the global burden of diseases, through both acute infection and their chronic sequelae. We identified an insertion-deletion variant thought to affect *NFKB1* expression as a causal variant—central to serological responses to diverse infectious agents, risk of infection, immune cell survival, antibody production, and inflammation.

A common *NFKB1* variant detected through antibody analysis in UK Biobank predicts risk of infection and allergy

Amanda Y. Chong,[1,*] Nicole Brenner,[2] Andres Jimenez-Kaufmann,[3] Adrian Cortes,[4] Michael Hill,[5] Thomas J. Littlejohns,[6] James J. Gilchrist,[1,7] Benjamin P. Fairfax,[8] Julian C. Knight,[1,9] Flavia Hodel,[10,11] Jacques Fellay,[10,11,12] Gil McVean,[4] Andres Moreno-Estrada,[3] Tim Waterboer,[2] Adrian V.S. Hill,[1,13] and Alexander J. Mentzer[1,4,14,*]

Summary

Infectious agents contribute significantly to the global burden of diseases through both acute infection and their chronic sequelae. We leveraged the UK Biobank to identify genetic loci that influence humoral immune response to multiple infections. From 45 genome-wide association studies in 9,611 participants from UK Biobank, we identified *NFKB1* as a locus associated with quantitative antibody responses to multiple pathogens, including those from the herpes, retro-, and polyoma-virus families. An insertion-deletion variant thought to affect *NFKB1* expression (rs28362491), was mapped as the likely causal variant and could play a key role in regulation of the immune response. Using 121 infection- and inflammation-related traits in 487,297 UK Biobank participants, we show that the deletion allele was associated with an increased risk of infection from diverse pathogens but had a protective effect against allergic disease. We propose that altered expression of *NFKB1*, as a result of the deletion, modulates hematopoietic pathways and likely impacts cell survival, antibody production, and inflammation. Taken together, we show that disruptions to the tightly regulated immune processes may tip the balance between exacerbated immune responses and allergy, or increased risk of infection and impaired resolution of inflammation.

Introduction

There is significant evidence that human genetic variation influences susceptibility to acute disease caused by infectious pathogens. Case-control genome-wide association studies (GWASs) of a variety of pathogens and infections, such as human immunodeficiency virus, hepatitis B and C viruses, leprosy, typhoidal and non-typhoidal *Salmonella*, meningococcal disease, and more recently SARS-CoV-2, have identified a range of variants spanning pathogen recognition, immune activation, and immunologic memory robustly associated with differential susceptibility.[1–10] Furthermore, using large biobanks of human genetic data linked with participant self-reporting, studies such as 23andMe are replicating and building on these findings from case-control analyses of infection susceptibility.[11] Together, these studies are helping improve our understanding of how genetic variants influence the delicate balance of immunity and risk of invasive infection that give rise to diseases such as meningitis, mononucleosis, or plantar warts and more autoimmune reactions to infections such as rheumatic heart disease. Despite these advances, there remain challenges to using both study approaches. Firstly, achieving sufficient power to discover genetic signals in case-control studies is limited by challenges in recruiting large numbers of cases, and both study approaches may be affected by pathogen heterogeneity that may have an impact on signatures of genetic association, while the self-report approach will be influenced by limitations in accuracy and recall of self-reporting.

Antibodies are a critical component of response to infection and represent a marker of prior exposure or chronic carriage with agents capable of latent or chronic infection states. Now that methods are available to accurately measure multiple antibodies simultaneously in large numbers of individuals, these represent stable biological markers that can be used to understand disease risk. Furthermore, there is evidence from vaccine studies that the magnitude of antibody response and levels of circulating antibody may reflect the likelihood of protection against either primary or subsequent disease, although this may depend on the antigen to which the antibody is directed, and the functional impact of antibody binding and activity.[12]

[1]The Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; [2]Division of Infections and Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; [3]Advanced Genomics Unit, National Laboratory of Genomics for Biodiversity (LANGEBIO), CINVESTAV, Irapuato, Mexico; [4]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; [5]MRC-Population Health Research Unit, University of Oxford, Oxford, UK; [6]Nuffield Department of Population Health, University of Oxford, Oxford, UK; [7]Department of Paediatrics, University of Oxford, Oxford, UK; [8]Department of Oncology, University of Oxford, Oxford, UK; [9]NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK; [10]Global Health Institute, School of Life Sciences, EPFL, Lausanne, Switzerland; [11]Swiss Institute of Bioinformatics, Lausanne, Switzerland; [12]Precision Medicine Unit, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland; [13]The Jenner Institute, University of Oxford, Oxford, UK
[14]Senior author
*Correspondence: amanda.chong@well.ox.ac.uk (A.Y.C.), alexander.mentzer@ndm.ox.ac.uk (A.J.M.)
https://doi.org/10.1016/j.ajhg.2023.12.013

Human genetic variation is recognized to significantly influence total antibody levels.[13] Genetic analyses of antibody response levels against vaccine preventable infections, such as hepatitis B virus (HBV), have demonstrated a strong influence of genetic loci, including the major histocompatibility complex (MHC) in predicting antibody magnitude overlapping with GWAS of disease susceptibility.[14] The MHC has also been implicated in varied antibody responses to infections, including JC virus and influenza.[15–19] Furthermore, primary immunodeficiencies, such as common variable immunodeficiency (CVID) and X-linked agammaglobulinemia, are well recognized disorders that increase the risk of multiple infections owing to impaired humoral function and antibody deficiencies.[20–22]

The availability of antibody data against 45 antigens representing 20 infectious agents in combination with genome-wide genotyping data in a subset of UK Biobank participants now offers the opportunity to investigate whether there are common variants that influence both response to antibodies and subsequent susceptibility to both acute infectious disease and their chronic sequelae. While previous studies have identified genetic associations with single antigens or pathogens using this UK Biobank dataset,[17,23] we hypothesize that there are genetic variants and loci that have a common effect on antibody responses to multiple infections through effects on central immunological pathways. To identify genetic loci that influence antibody titers against multiple infections, we performed GWASs for a magnitude of antibody responses to each antigen and identified *NFKB1* as a distinct locus of interest in contrast to the other, more familiar loci that were identified, including immunoglobulin genes, the MHC, and *FUT2*. Given the role of *NFKB1* in inflammation and the induction and regulation of immune responses to infection, we further investigated the association between *NFKB1* variation and infectious and inflammatory diseases using data from 487,297 individuals. Finally, to better understand the impact of variation in this locus on underlying immune mechanisms, we explored the effect of this variant on blood cell traits and cellular expression data both within UK Biobank and in dedicated expression quantitative trait datasets.

Subjects and methods

Serology and GWAS

We used serological measurements of infectious agents released as part of the UK Biobank Data Showcase[24,25] to investigate genetic loci associated with susceptibility to infectious disease. This data consist of measured antibody responses against 45 antigens from 20 viral, bacterial, and protozoan pathogens present in the UK (Tables S1–S3) from a randomly selected subset (n = 9,695) of the UK Biobank cohort. Viral pathogens included on the panel included herpes simplex virus types 1 and 2 (HSV-1 and -2), varicella-zoster virus (VZV), Epstein-Barr virus (EBV), human cytomegalovirus (CMV), human herpesvirus 6 and 7 (HHV-6 and -7), Kaposi's sarcoma-associated herpesvirus (KSHV), HBV, hepatitis C virus (HCV), human T-lymphotropic virus 1 (HTLV-1), human immunodeficiency virus 1 (HIV-1), BK and JC polyomaviruses (BKV and JCV), Merkel cell polyomavirus (MCV), and human papillomavirus 16 and 18 (HPV-16 and -18). Other pathogens assayed were the bacterial pathogens *Chlamydia trachomatis* and *Helicobacter pylori*, as well as protozoan parasite *Toxoplasma gondii*. Target antigens were selected based on known biological function and validation assays (Table S1). Measurements from samples collected at the initial assessment visit were used for analysis. Mean age of participants was 56.55 years at time of sample collection.

For all association tests, we used the genotyped and imputed autosomal variant datasets available for the UK Biobank cohort.[26] This consists of 805,426 high-quality genotyped variants, generated from the UK BiLEVE and UK Biobank Axiom arrays, and 93,095,623 imputed autosomal variants were imputed via the Haplotype Reference Consortium (HRC), UK10K, and 1000 Genomes phase 3 reference panels.[27–30] Quality control filters for poorly performing variants and samples prior to imputation are described in Bycroft et al.[26] UK Biobank has obtained ethics approval from the North West Multi-centre Research Ethics Committee (approval number: 11/NW/0382) and had obtained informed consent from all participants.

For our analysis we had 9,611 unrelated individuals that had genetic and serological data available. Antibody responses across all participants regardless of serostatus were normalized using a rank-based inverse normal transformation (INT). All estimated betas are reported relative to standard deviations of the normalized distribution. We measured the Spearman's rank correlation between traits to determine whether there was any evidence of correlation between antibody traits (Figure S1). Correlation measurements were repeated after regressing out the effects of age, sex, and lead genetic variants as identified from GWAS analyses (Figure S2). Regression analyses were carried out using fitted generalized linear models as implemented in the python statsmodels package and normalized MFI values.

Association analyses using imputed genotype data and both the untransformed and normalized data from quantitative serology traits were carried out using linear mixed models as implemented in BOLT-LMM (minor allele frequency [MAF] >0.01, imputation infoscore >0.3;[31]) including age at recruitment and genetic sex as covariates. All participants were included irrespective of ethnic origin as the use of linear mixed models allows us to account for population substructure through the incorporation of a genetic relatedness matrix (GRM) as a random effect covariate. Reported betas and standard errors are based on normalized distributions and reflect magnitude of response. To identify the genomic regions and candidate genes associated with antibody response to each of the tested antigens, we first identified SNPs associated with tested antigens based on genome-wide significance ($p < 5 \times 10^{-8}$) and suggestive significance ($p < 1 \times 10^{-5}$). Starting with the most significant SNPs (peak SNP), we defined associated regions for each antigen based on calculated linkage disequilibrium (LD) from best-guess genotypes ($R^2 \geq 0.2$, calculated with PLINK; Chang et al.[32]) between the peak SNP and other SNPs reaching suggestive significance for the dataset. This was repeated for each SNP that was not linked with a previous peak SNP. We then merged overlapping regions across all traits to define independent regions of association for further analysis.

Fine mapping of GWAS signals

For regions containing SNPs with an association p value less than the Bonferroni corrected p value (p < 1.1 × 10^{-9}), we then performed conditional analyses to determine if there were multiple independent signals present within the region. Analyses were carried out in GCTA using the –mlma option and adjusting for allelic dosage of the top SNP for each associated trait.[33] These gene dosages were also used to test for interdependence between these associated regions and other regions of interest.

FINEMAP was used (default settings, R2 = 0.8; Benner et al.[34]) to identify potential causal groups (Bayes factor ≥2) of SNPs within each of these regions based on summary statistics for each of the associated traits. These were then compared across antigens and merged to define haplotypes of SNPs based on shared SNPs and direction of effect. Causal groups sharing SNPs were only merged if the direction of effect across all antigens were in a consistent direction relative to the minor allele.

Potential causal SNPs were identified from haplotypes shared across multiple antigens. To do this, we first checked the location of SNPs relative to the Ensembl GRCh37.p13 gene annotations to identify the nearest genes. The predicted effects of each variant were then retrieved using the Ensembl Variant Effect Predictor.[33]

Multivariate GWAS

Following fine mapping of association signals in the *NFKB1* region, we carried out a multivariate GWAS using the seven phenotypes with a predicted causal SNP in *NFKB1*. The phenotypes included were EBV (EAD), HHV-6 (IE1A), HHV-7 (U14), HIV (env), HPV-18 (L1), KSHV (K8.1), and *T. gondii* (sag1). A multivariate GWAS was carried out using SNPTEST (v2.5.6;[35]), and the INT normalized phenotypes. Age at recruitment, sex, and the top four principal components were included as covariates, as SNPTEST does not fit a GRM as part of the association model. Principal components were calculated from LD-pruned genotype data using PLINK (v1.90b3).

Cohorte Lausannoise replication study and meta-analysis

The Cohorte Lausannoise (CoLaus/PsyCoLaus) study includes 6,188 individuals of European ancestry living in Lausanne, Switzerland who were randomly selected from the general population. Recruitment and sampling protocols are described by Firmann et al.[36] The institutional ethics committee of the University of Lausanne, which afterward became the Ethics Commission of Canton Vaud (www.cer-vd.ch), approved the baseline CoLaus study (ref. 16/03, decisions of January 13 and February 10, 2003), and written consent was obtained from all participants.

Genotyping data for 5,399 individuals were obtained using the Affymetrix Axiom Biobank array. Following quality control steps to remove poorly performing variants and samples, imputed genotypes were generated with the Sanger Imputation Service.[30] Genotypes were phased using EAGLE2 (v2.0.5; Loh et al.[29]) and the HRC panel,[30] and imputation was carried out using the HRC reference panel, 1000 Genomes Phase 3 data, and the UK10K reference panel.[27,28] Serology phenotypes were generated using a similar Multiplex Serology Panel to the UK Biobank data as described above (Table S3).[16,25]

Genome-wide association analyses were carried out using 4,216 individuals with imputed genotype data and serology data and linear mixed models as implemented in GCTA (1.91.3beta) and thresholded high-quality imputed genotype data (imputation in-foscore >0.8, MAF >0.01, $P_{HWE} < 10 \times 10^{-7}$). Age, sex, and the top three principal components were included as covariates. For antigens with a signal at *NFKB1* and that were also available from the CoLaus/PsyCoLaus cohort, we performed GWASs of antibody levels that were common across the two cohorts and extracted the set of intersecting SNPs for meta-analysis using Metasoft (v2.0.1; Han and Eskin[37]).

Infection and inflammation associations in UK Biobank

We used self-reported non-cancer illness (Field ID: 20002) and prevalent and incident hospital inpatient International Classification of Disease version 10 (ICD-10) codes (Field ID: 41270) to define cases and controls for further infection- and inflammation-related phenotypes within UK Biobank. For self-reported conditions, cases and controls were drawn from participants who reported the presence or absence of non-cancer-related conditions and had genetic data available, excluding those who did not respond to the question. Controls for infection-related traits derived from self-reported data were defined as individuals who had reported no major infections for infection-related phenotypes with genetic data available. Controls for other immune-related traits were drawn from all unaffected individuals with genetic data.

Similarly, potential cases and controls based on ICD-10 data were drawn from those individuals with hospital inpatient diagnoses and genetic data available. Case definitions are outlined in Tables S4 and S5. Controls for ICD-10-derived phenotypes were drawn from individuals with hospital inpatient and genetic data who were not affected by the trait of interest. In the case of infection-related traits, the selection was further restricted to those who had not reported a major infection or had no ICD codes relating to infection (e.g., Chapter I [A00-B99], Chapter X: Diseases of the respiratory system: Acute upper respiratory infections [J00-J06], Influenza and pneumonia [J09-J18], other acute lower respiratory infections [J20-J22]).

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

Association tests for these traits were carried out using linear mixed models as implemented in SAIGE (v 0.36.3.2; Zhou et al.[38]), also including age at recruitment and genetic sex as covariates, except for traits where only one sex was present in the cases, in which case sex was not included, but controls were drawn only from individuals of that sex (Tables S4 and S5 for more information). Due to relatively small numbers of individuals for which cause of death was available (approximately 25,000), we used logistic regression to assess the relationship between rs28362491:delATTG and the contribution of infection or inflammatory conditions to cause of death. Similar to the GWAS analysis described above, we used imputed dosages of rs28362491:delATTG with age and sex as covariates. We also restricted analysis to individuals in the White British subset as defined by UK Biobank.

Association with blood cell counts

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

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

Expression quantitative trait locus analysis

To assess the regulatory function of rs28362491, we used the tightly linked 5′UTR SNP rs72696119 ($r^2 = 0.999$, D′ = 0.999). We correlated rs72696119 genotype with *NFKB1* mRNA expression in naive and stimulated primary immune cell subsets from healthy European adults, using data from expression quantitative trait locus (eQTL) studies. $CD56^+CD3^-$ NK cells, $CD19^+$ B cells, $CD14^+$ monocytes, and $CD16^+$ neutrophils were separated, genotyped, and gene expression quantified as previously described.[39–42] Following quality control and normalization of gene-expression data, we correlated rs72696119 genotype with *NFKB1* RNA expression in each cell type: B cells (n = 279), NK cells (n = 245), neutrophils (n = 101), naive monocytes (n = 414), and stimulated monocytes (lipopolysaccharide [LPS] 2 h, n = 261; LPS 24 h, n = 322; interferon gamma [IFNγ] 24 h, n = 367). *NFKB1* expression was correlated with genotype by linear regression and analysis of variance (ANOVA), including the first 25 principal components of gene-expression data in each cell type/condition to account for confounding variation. p- values are calculated with F-tests (1 degree of freedom). Statistical analysis was performed in R. We used publicly available eQTL data from the DICE database for $CD4^+$ and $CD8^+$ T cells to examine association of rs72696119 with *NFKB1* expression in naive T cells (Schmiedel et al.[43]; https://dice-database.org/).

Colocalization of eQTL and GWAS signals

Colocalization testing between eQTL signals and antibody and blood cell GWAS association signals was carried out using coloc (v5.0.1; Wallace[44]) with susieR (v0.11.92; Wang et al. and Zou et al.[45,46]) and the intersection of variants from all GWAS phenotypes and the eQTL traits from the region immediately upstream of and across NFKB1. LD was calculated independently for each dataset using PLINK. Colocalization was tested between pairs of antibody traits: between HHV6 (IE1A) and blood cell traits, between HHV6 (IE1A) and eQTL datasets, and between related blood cell traits and eQTL datasets (e.g., monocyte count and unstimulated and stimulated monocyte eQTL datasets) (Table S13).

Mendelian randomization

The causal effect of blood cell counts on antibody response and risk of disease were tested using one- and two-sample Mendelian randomization (MR) by using rs28362491 as the genetic instrumental variable. We used normalized cell counts for RBCs, basophils, eosinophils, neutrophils, monocytes, lymphocytes, and platelets. Antibody response measures and health-record-derived traits were restricted to those with a significant association at rs28362491 (antibody response: EBV [EAD and ZEBRA], HHV-6 [IE1A], HHV-7 [U14], HPV-18 [L1], HTLV-1 [env], KSHV [K8.1], and *T. gondii* [sag1]; health records: hay fever, influenza [ICD], *Legionella* spp. [ICD], *Neisseria meningitidis* [ICD], papillomavirus [ICD], psoriasis, ulcerative colitis, Alzheimer's disease, cervical intraepithelial neoplasia [SR], allergy, and infection). Bias as a result of overlapping samples between datasets for two-sample analyses are expected to be negligible given the small size of the antibody dataset in comparison with the full UK Biobank cohort.[47]

We used the inverse-variance weighted (IVW) method as implemented in the R package MendelianRandomization[48,49] for two-sample MR and multiplicative structural mean model (MSMM; binary outcome) from R package OnesampleMR (https://github.com/remlapmot/OneSampleMR) for one-sample analyses. Power calculations for pairs of traits were carried out using an online power calculator (https://sb452.shinyapps.io/power/). The variance in the exposure explained by rs28362491 (R^2) and F statistics for instrument strength was also calculated. An F statistic >10 was taken to be indicative of a low risk of weak instrument bias in MR analyses.

Results

We have previously described the use of a Multiplex Serology panel to measure immunoglobulin G (IgG) antibody responses against 45 antigens from 20 infectious agents implicated in the pathogenesis of chronic disease, applied to 9,695 individuals in UK Biobank.[24] We observed that antibody responses to antigens from the same pathogen were more strongly correlated with each other (Spearman's rho 0.10–0.86; Figure S1) than with those of other pathogens (Spearman's rho −0.05-0.75) and that all antibody responses were generally weakly correlated with each other with the exception of KSHV (K8.1), HIV (env), HTLV-1 (env), and HPV-18 (L1), which were all strongly correlated (Spearman's rho 0.70–0.75). However, seroprevalence estimates for these pathogens in UK Biobank are very low (0.2%–8.1%; Mentzer et al.[24]). Since cross-reactivity between antigens was found to be negligible for all antigens included in the current panel, except for HTLV-1 gag,[24] it is likely that shared genetic variants and environmental factors influence these correlated responses. Using the 9,611 individuals with both serology data and imputed genetic data, we performed GWASs of quantitative antibody responses to 45 antigens to identify these common genetic loci that influence response to multiple infections.

We identified 27 genome-wide significant (p < 5 × 10^{-8}) loci associated with magnitude of antibody responses to any antigen (Figure 1; Table S6). We observed four genomic regions, localizing to *NFKB1* (most significant association with HHV-6 [IE1A]; beta = −0.10; p = 1.30 × 10^{-10}), the extended MHC region (EBV [EBNA-1]; beta = −0.32; p = 6.90 × 10^{-97}), the IGH locus (HPV-16 [E6]; beta = 0.19; p = 1.60 × 10^{-10}), and *FUT2* (JC polyomavirus [VP1]; beta = −0.13; p = 1.70 × 10^{-21}), which were associated with antibody response to multiple antigens and where at least one association signal passed Bonferroni correction (p < 1.1 × 10^{-9}). All four genetic regions are well recognized to drive differences in infection susceptibility.[10,11,14–17,23,50]

Associations were observed within the extended MHC region, centered mainly around human leukocyte antigen (HLA) genes, for at least one antigen from each of the 20 pathogens, with associations for human CMV, *Chlamydia trachomatis*, EBV, HHV-6, HHV-7, HPV-16, HSV-1 and -2, HTLV-1, JCV, MCV, *Toxoplasma gondii* (*T. gondii*), and VZV reaching genome-wide significance (Table S7). This suggests that, with adequate power, we are now able to resolve HLA associations with responses to antigens considered
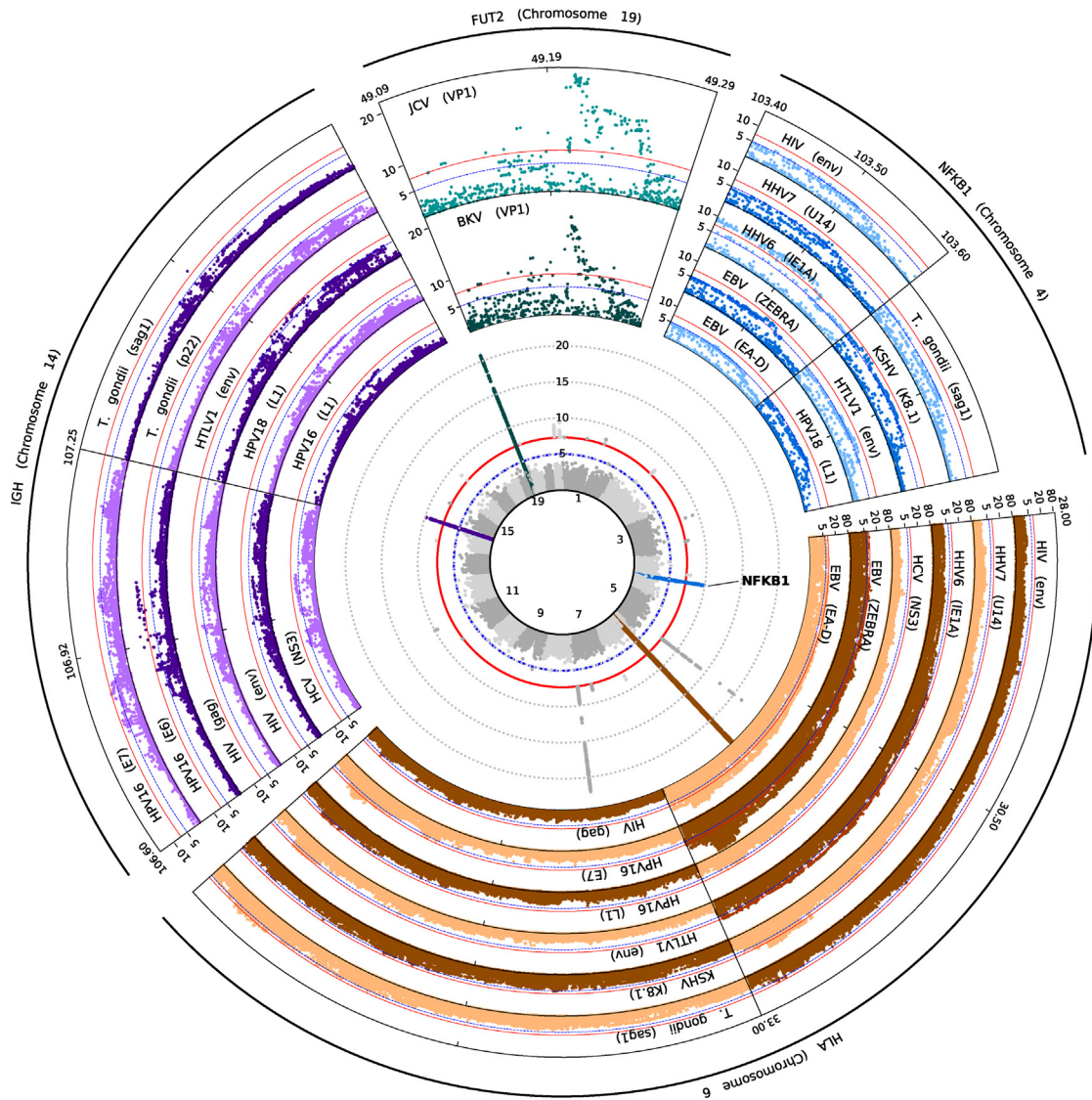
Figure 1. *NFKB1* is a key locus affecting antibody response and risk of infection and inflammatory disease
Central inset: stacked Manhattan plot of association statistics for antibody response to 45 pathogen-derived antigens. Signals for the four loci with multiple associations are colored: *NFKB1* (blue), the MHC locus (brown), the IGH locus (purple), and *FUT2* (teal). Surrounding insets: Regional association plots for specific antigens demonstrating signals at each of the four highlighted loci. Y axis (all plots) is –log$_{10}$(P).

most clinically relevant (such as EBNA-1 and EA for EBV; Figure 1), such as those described from the same UK Biobank cohort.[16,17,23]

NFKB1 is associated with antibody responses against multiple pathogens

Of particular interest is the identification of multiple variants in *NFKB1* as a common locus associated with antibody responses across multiple pathogens and pathogen types, including EBV, HPV-18, HIV, HTLV-1, KSHV, and *T. gondii*. To confirm the likely contribution of this locus to multiple infectious antigen responses, we replicated this association, demonstrating a consistent direction of effect for three antigens (EBV [EA-D] beta = −0.12, p = 9.15 × 10^{-7}; EBV [ZEBRA] beta = −0.12, p = 7.20 ×

10^{-7}; and HHV7 [U14] beta = −0.06, p = 1.81 × 10^{-2}) using an independently recruited cohort (CoLaus/PsyCoLaus[16]) where data using the same Multiplex Serology technology was available but with a different set of antigens (Tables 1 and S8).

Using a combination of fine mapping and examination of possible variant effects, we identified the most-likely causative variant to be rs28362491, a 4 bp insertion-deletion (indel) variant within a promoter region upstream of *NFKB1*. The rs28362491 deletion is proposed to disrupt a promoter binding site, lowering *NFKB1* expression.[51] We found rs28362491 to be part of a cluster of tightly linked SNPs (R^2 > 0.8) predicted to contain at least one causal variant for 7 antigens: EBV (EAD), HHV-6 (IE1A), HHV-7 (U14), HPV-18 (L1), HIV (env), KSHV (K8.1), and *T. gondii*

Table 1. UK Biobank and CoLaus/PsyCoLaus meta-analysis results for rs28362491

Phenotype	UK Biobank (n = 9611)				CoLaus/PsyCoLaus (n = 4216)				Meta-analysis		
	Seropositive	beta	SE	P	Seropositive	beta	SE	P	beta	SE	P
EBV (EAD)	8,278	−0.059	0.014	3.20×10^{-5}	3,257	−0.117	0.024	9.15×10^{-7}	−0.074	0.012	1.13×10^{-9}
EBV (ZEBRA)	8,763	−0.062	0.014	1.30×10^{-5}	3,718	−0.116	0.023	7.20×10^{-7}	−0.076	0.012	3.01×10^{-10}
HHV6 (IE1A)	7,469	−0.086	0.015	2.90×10^{-9}	766	−0.03	0.024	2.03×10^{-1}	−0.071	0.012	1.03×10^{-8}
HHV7 (U14)	9,103	−0.057	0.014	5.00×10^{-5}	2,244	−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)	2,190	−0.063	0.014	1.20×10^{-5}	1,328	−0.004	0.023	8.73×10^{-1}	−0.047	0.012	1.33×10^{-4}

(sag1). A Bayesian-based multivariate GWAS using these phenotypes confirmed the presence of a shared association across *NFKB1* (Figure S4), with rs28362491 also demonstrating a significant association (log10 Bayes factor = 5.056), as well as shared associations across the HLA region (rs9272339; log10 Bayes factor = 25.976).

NFKB1 is one of five genes that encode the NF-κB family of transcription factors. The NF-κB family plays a critical role in induction and mediation of pro-inflammatory response to infection and for development and maintenance of blood cells and immune tissues through modulation of apoptosis. Transcriptomic profiling of immune response to pathogens suggests that *NFKB1* forms part of a larger set of genes involved in regulation of immune response to a wide range of pathogen infections.[52] Furthermore, mutations in *NFKB1* have been identified as one of the most common causes of monogenic CVID in European populations.[20,22]

In addition, rs28362491 has been associated with multiple disease phenotypes in a range of previously published genetic association studies. Examples of such traits with significant evidence of effect include coronary artery disease with consistent reports of increased odds of disease with presence of the deletion demonstrated in European[53] (odds ratio [OR] = 2.88 [95% confidence interval (CI) 1.21–6.84]; Figure 2; Table S9), Indian[54] (OR = 1.26 [1.03–1.55]), and Uyghur[55] (OR = 1.58 [1.22–2.05]) populations; risk of lung injury in acute respiratory distress syndrome in Europeans[56] (OR = 3.7 [1.8–7.9]); and ulcerative colitis in Europeans[51,57] (OR = 1.57 [1.14–2.16]).
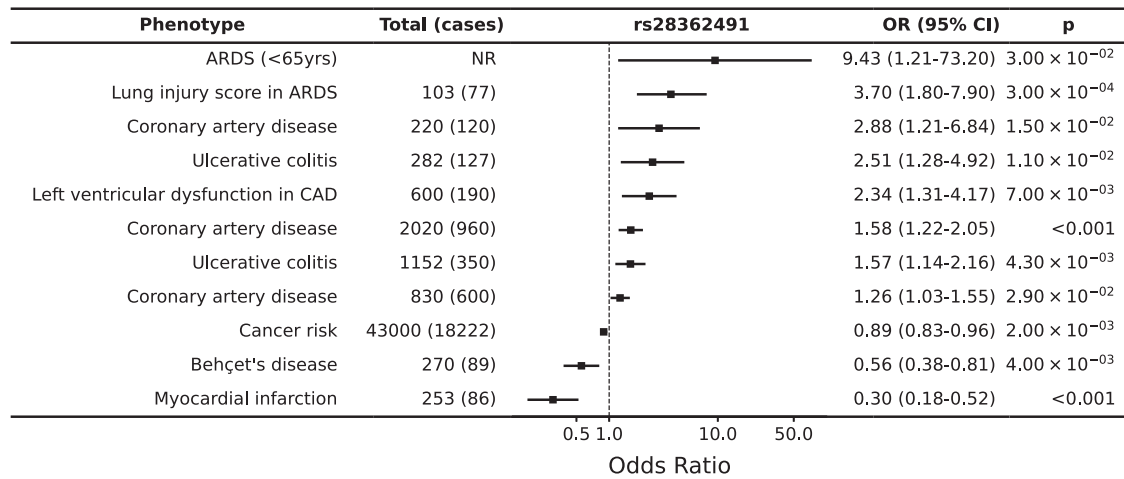
The same region has also been found to be associated in the same direction for other traits where, although the rs28362491 variant was not reported, it is likely that the effect may co-localize. These traits include tonsillectomy[11] (OR = 1.08 [1.06–1.08]; rs230523, R^2 = 0.759), ankylosing spondylitis[58] (OR = 1.12 [1.08–1.16]; rs3774937, R^2 = 0.751), primary biliary cholangitis[59] (OR = 1.34 [1.11–1.23]; rs230534, R^2 = 0.745), and mouth ulcers[60] (OR = 1.03 [1.02–1.04]; rs4699030, R^2 = 0.863) among others (Table S9). Intriguingly, the same locus and/or variant was associated with other traits in the opposite direction, including allergic rhinitis[61] (OR = 0.96 [0.95–0.97]; rs12509403, R^2 = 0.556), hay fever and eczema[62] (OR =

0.96 [0.95–0.97]; rs230507, R^2 = 0.776), and cancer risk[63] (OR = 0.89 [0.83–0.96]; rs28362491).

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

We found that the direction of effect for rs28362491 was consistent across 9 antigens for which we observed a signal for serum antibody responses in UK Biobank (EBV [EAD and ZEBRA], HHV-6 [IE1A], HHV-7 [U14], HPV-18 [L1], HIV [env], HTLV-1 [env], KSHV [K8.1], and *T. gondii* [sag1]), with the deletion allele (rs28362491:delATTG) being associated with lower antibody responses in UK Biobank (Figure 3). Given this observation, with our observed replication in CoLaus/PsyCoLaus and the published associations with multiple infectious or inflammatory conditions (Table S9), we further tested for the effect of rs28362491:delATTG specifically on susceptibility to infection or inflammatory disease in UK Biobank, using prevalent and incident hospital inpatient (ICD-10) or self-reported cases of infection and inflammatory conditions in 487,297 participants. While neither hospital inpatient data nor self-reporting are likely to provide a comprehensive record of infection or pathogen exposure, they provide a valuable counterpoint to help understand the relationship between antibody titers and disease.

Firstly, we replicated the previously published associations with diverse traits such as allergic rhinitis[61] and ulcerative colitis,[51,57,58,64] confirming the variant's role in a variety of traits (Figure 3; Tables S9 and S10). Then we explored the effect of the variant on reported traits related to the measured antibody responses. For example, we found that although rs28362491:delATTG was consistently associated with reduced antibody levels against all measured HPV antigens, there was an increased odds of reporting cervical intraepithelial neoplasia (OR = 1.17 [1.02–1.18]) or having an ICD code of papillomavirus (OR = 1.09 [1.02–1.34]) (Figure 3). We observed a similar pattern for EBV where the deletion was associated with reduced responses against multiple EBV antigens (e.g., EA-D: OR = 0.94 [0.92–0.97]), yet there was a trend toward increased odds of self-report of EBV infection/infectious mononucleosis (OR = 1.07 [0.98–1.18]) and multiple sclerosis (which

Phenotype	Total (cases)	rs28362491	OR (95% CI)	p
ARDS (<65yrs)	NR		9.43 (1.21-73.20)	3.00×10^{-02}
Lung injury score in ARDS	103 (77)		3.70 (1.80-7.90)	3.00×10^{-04}
Coronary artery disease	220 (120)		2.88 (1.21-6.84)	1.50×10^{-02}
Ulcerative colitis	282 (127)		2.51 (1.28-4.92)	1.10×10^{-02}
Left ventricular dysfunction in CAD	600 (190)		2.34 (1.31-4.17)	7.00×10^{-03}
Coronary artery disease	2020 (960)		1.58 (1.22-2.05)	<0.001
Ulcerative colitis	1152 (350)		1.57 (1.14-2.16)	4.30×10^{-03}
Coronary artery disease	830 (600)		1.26 (1.03-1.55)	2.90×10^{-02}
Cancer risk	43000 (18222)		0.89 (0.83-0.96)	2.00×10^{-03}
Behçet's disease	270 (89)		0.56 (0.38-0.81)	4.00×10^{-03}
Myocardial infarction	253 (86)		0.30 (0.18-0.52)	<0.001

Odds Ratio

Figure 2. rs28362491:delATTG is associated with increased risk of a range of disease associated with dysregulation of inflammation
References are listed in Table S9. ARDS, acute respiratory distress syndrome; CAD, coronary artery disease.

has been plausibly causally linked with EBV infection; OR = 1.05 [0.98–1.11] in our UK Biobank data), as well as a similar effect for *H. pylori*.

We then proceeded to test the association of the variant with a wider number of infectious and inflammatory diagnoses. Out of 39 infections tested, the deletion allele increased risk of infection with influenza (OR = 1.120 [1.000–1.253]; p = 4.90×10^{-2}), legionella (OR = 1.524 [1.097–2.118]; p = 1.21×10^{-2}), and papillomavirus (OR = 1.095 [1.019–1.177]; p = 1.38×10^{-2}) and decreased the risk of infection with *Neisseria meningitidis* (OR = 0.704 [0.509–0.974]; p = 3.43×10^{-2}). Since many infections were limited in number, we also tested the effect of the deletion allele against a phenotype consisting of any infection and found that overall, there was a significant increase in risk associated with the deletion (OR = 1.014 [1.001–1.027], p = 3.31×10^{-2}; Table S10). Furthermore, we found that the same variant was associated with an increased risk of diseases associated with dysregulation of inflammatory responses and chronic inflammation, such as psoriasis (OR = 1.044 [1.005–1.083]; p = 2.55×10^{-2}). In contrast, we found that the deletion allele was protective against allergic diseases, such as allergic rhinitis (OR = 0.965 [0.949–0.981]; p = 3.57×10^{-5}), which are associated with acute inflammatory responses.

Altogether, these results demonstrate that this variant is associated with diverse groups of infectious or allergic diagnoses. It appears that the deletion tends to increase the risk of infectious or chronic inflammatory diagnoses possibly associated with reduced antibody responses. Conversely, the insertion is associated with increased antibody response (and thus reduced risk of infectious diagnoses) but will come at an increased risk of allergic conditions. Given these observed divergent patterns of association with risk of inflammatory and infectious or allergic disease, we proceeded to test for risk of mortality in UK Biobank participants due to allergic, infectious, or inflammatory disease. Although low in frequency given the relative
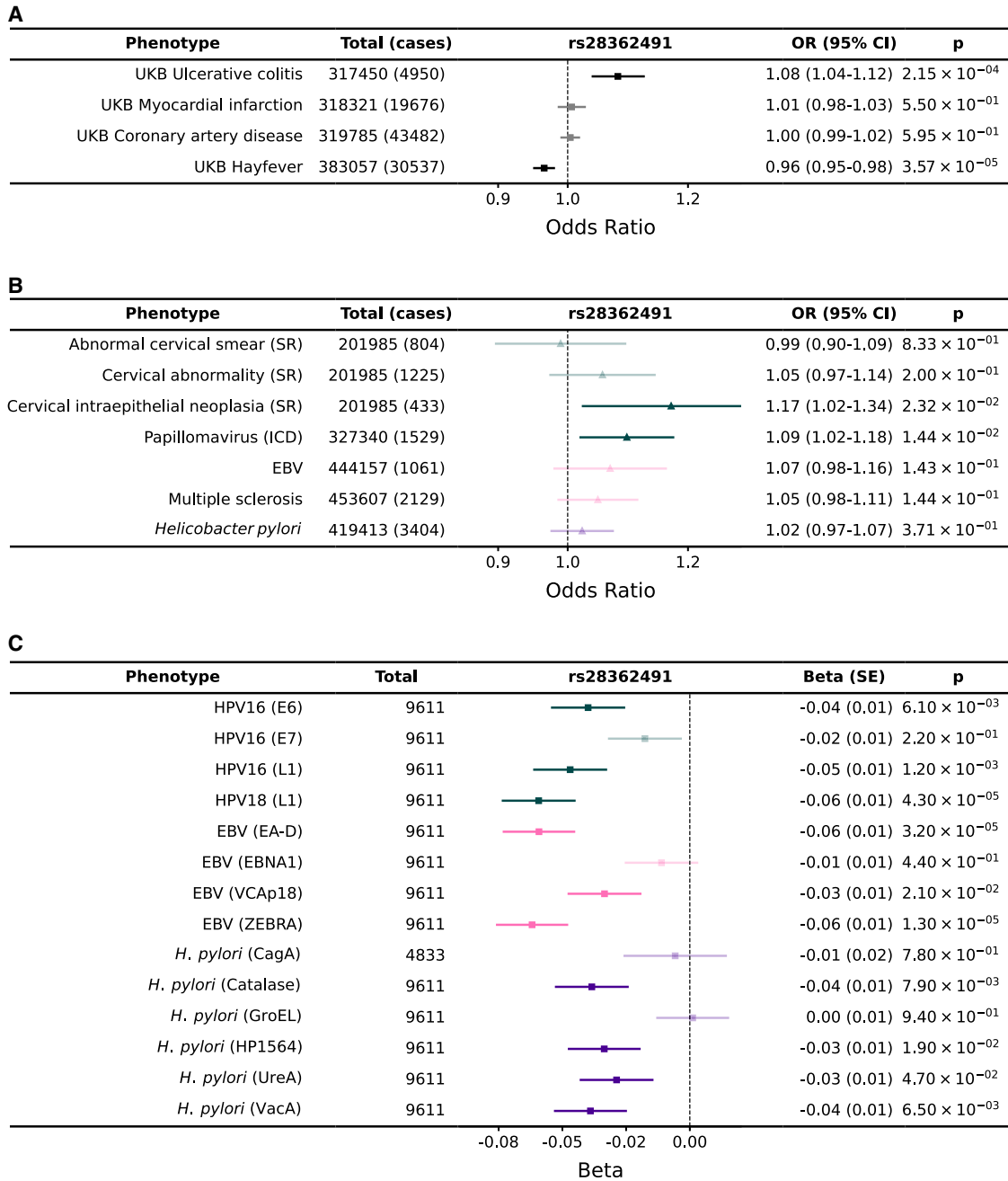
immaturity of the UK Biobank cohort study, we observed similar trends of a protective effect of the deletion variant against death from allergic disease (OR = 0.846 [0.658–1.088], p = 0.192) and an increased risk of infection (OR = 1.065 [0.982–1.155], p = 0.129) and inflammation (OR = 1.038 [0.991–1.087], p = 0.117) (Table S11).

NFKB1 influences hematopoiesis and immune cell function

Having found significant evidence of the rs28362491:delATTG variant being associated with phenotypic disease endpoints and our intermediate antibody traits, we next sought to use these associations to understand the likely molecular pathways disrupted leading to increased infection risk. We first investigated the effect of rs28362491:delATTG on WBC counts and function (Table S12). Using cell-count data from UK Biobank, we observed that the deletion allele was not significantly associated with lower overall WBC counts (p = 0.96), but instead, carriers of the deletion allele had a higher total number of lymphocytes (beta = 1.98×10^{-2}, p = 3.30×10^{-24}) and eosinophils (beta = 6.49×10^{-3}, p = 9.30×10^{-4}) than those without. Individuals with the deletion also had lower counts of basophils (beta = -1.25×10^{-2}, p = 2.00×10^{-9}), monocytes (beta = -1.50×10^{-2}, p = 2.70×10^{-15}), neutrophils (beta = -7.27×10^{-3}, p = 2.50×10^{-4}), and platelets (beta = -5.79×10^{-3}, p = 1.30×10^{-3}) (Figure 4).

We next interrogated the differential expression of *NFKB1* in diverse blood cell types and found that the effect varied depending on the stimulation state of the cell. As rs28362491 is upstream of *NFKB1* and is not captured in the eQTL datasets, we used a tightly linked SNP (rs72696119, $r^2 = 0.999$ D' = 0.999) in the 5′ untranslated region (UTR) of *NFKB1* as a tagging SNP for eQTL analyses. We found that rs72696119 is a *cis*-eQTL for *NFKB1*, associated with increased expression in resting monocytes (p = 1.20×10^{-2}), neutrophils (p = 8.00×10^{-3}), NK cells

A

Phenotype	Total (cases)	rs28362491	OR (95% CI)	p
UKB Ulcerative colitis	317450 (4950)		1.08 (1.04-1.12)	2.15×10^{-04}
UKB Myocardial infarction	318321 (19676)		1.01 (0.98-1.03)	5.50×10^{-01}
UKB Coronary artery disease	319785 (43482)		1.00 (0.99-1.02)	5.95×10^{-01}
UKB Hayfever	383057 (30537)		0.96 (0.95-0.98)	3.57×10^{-05}

Odds Ratio (0.9 – 1.0 – 1.2)

B

Phenotype	Total (cases)	rs28362491	OR (95% CI)	p
Abnormal cervical smear (SR)	201985 (804)		0.99 (0.90-1.09)	8.33×10^{-01}
Cervical abnormality (SR)	201985 (1225)		1.05 (0.97-1.14)	2.00×10^{-01}
Cervical intraepithelial neoplasia (SR)	201985 (433)		1.17 (1.02-1.34)	2.32×10^{-02}
Papillomavirus (ICD)	327340 (1529)		1.09 (1.02-1.18)	1.44×10^{-02}
EBV	444157 (1061)		1.07 (0.98-1.16)	1.43×10^{-01}
Multiple sclerosis	453607 (2129)		1.05 (0.98-1.11)	1.44×10^{-01}
Helicobacter pylori	419413 (3404)		1.02 (0.97-1.07)	3.71×10^{-01}

Odds Ratio (0.9 – 1.0 – 1.2)

C

Phenotype	Total	rs28362491	Beta (SE)	p
HPV16 (E6)	9611		-0.04 (0.01)	6.10×10^{-03}
HPV16 (E7)	9611		-0.02 (0.01)	2.20×10^{-01}
HPV16 (L1)	9611		-0.05 (0.01)	1.20×10^{-03}
HPV18 (L1)	9611		-0.06 (0.01)	4.30×10^{-05}
EBV (EA-D)	9611		-0.06 (0.01)	3.20×10^{-05}
EBV (EBNA1)	9611		-0.01 (0.01)	4.40×10^{-01}
EBV (VCAp18)	9611		-0.03 (0.01)	2.10×10^{-02}
EBV (ZEBRA)	9611		-0.06 (0.01)	1.30×10^{-05}
H. pylori (CagA)	4833		-0.01 (0.02)	7.80×10^{-01}
H. pylori (Catalase)	9611		-0.04 (0.01)	7.90×10^{-03}
H. pylori (GroEL)	9611		0.00 (0.01)	9.40×10^{-01}
H. pylori (HP1564)	9611		-0.03 (0.01)	1.90×10^{-02}
H. pylori (UreA)	9611		-0.03 (0.01)	4.70×10^{-02}
H. pylori (VacA)	9611		-0.04 (0.01)	6.50×10^{-03}

Beta (-0.08 – -0.05 – -0.02 – 0.00)

Figure 3. rs28362491:delATTG is associated with increased risk of a range of infections and disease associated with dysregulation of inflammation and protective against exacerbated immune responses and allergy
(A) Replication of published associations between rs28362491 and disease in UK Biobank.
(B) Odds ratios and 95% confidence intervals for association between rs28362491 and 7 health-record-derived disease traits derived from health record data (ICD-10 and self-report).
(C) Regression coefficients (betas) for association between rs28362491 and normalized antibody response against 14 antigens with equivalent ICD-10 or self-reported phenotypes in (B). Phenotypes in (B) and (C) are colored by pathogen: papillomavirus (teal), EBV (pink), and *H. pylori* (purple). Solid colors indicate traits with a p < 0.05. ICD, phenotype is derived from ICD-10 health record data; SR, phenotype is derived from self-reported health record data.

($p = 9.23 \times 10^{-4}$), and decreased expression in naive $CD4^{+}$ and $CD8^{+}$ T cells ($p = 3.50 \times 10^{-4}$ and $p = 1.40 \times 10^{-3}$) but not associated with differential expression in B cells ($p = 0.99$). All significant loci in all tested analyses were found to colocalize (e.g., HHV-6 [IE1A] and lymphocyte count: $PP_{(H4)} = 1.45 \times 10^{-7}$; Table S13).

As blood cell counts are available for the full UK Biobank cohort, we used Mendelian randomization analysis to test for causal links between the rs28362491:delATTG variant, blood cell traits, and measured antibody responses. To maximize power for these associations, given the small degree of sample overlap relative to the total number of
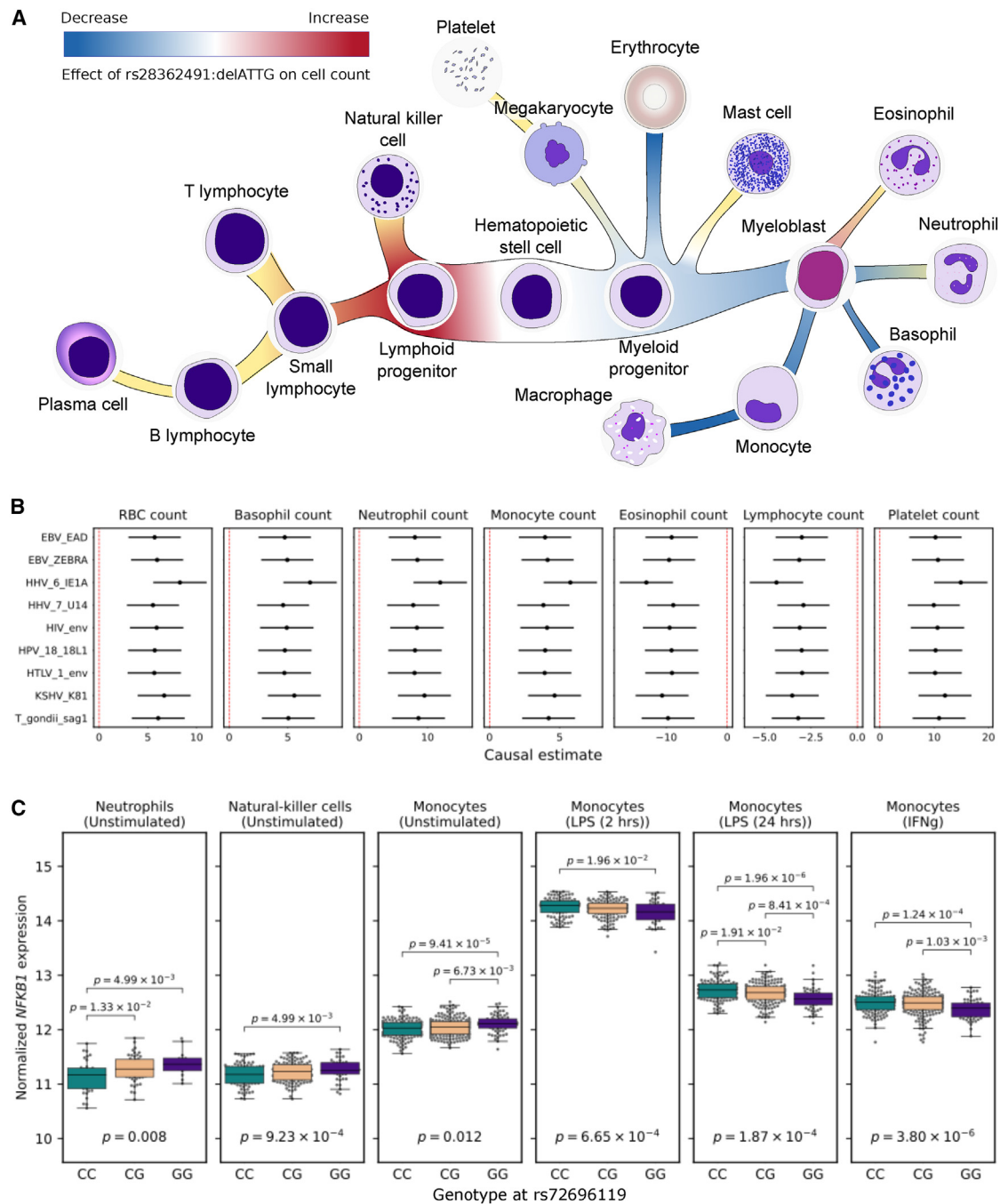
Figure 4. rs28362491 variation is also associated with differential hematopoiesis and *NFKB1* expression in white blood cells
(A) Effect of rs28362491:delATTG on WBC counts, indicated by red-blue shading. Yellow shaded regions indicate a lack of information on cell counts (mast cells, B and T cells) or signals that did not reach genome-wide significance (NK cells, platelets, neutrophils, and eosinophils).
(B) Causal estimates from two-sample Mendelian randomization analyses examining the effect of normalized blood cell counts on antibody responses.
(C) Differential expression by genotype at tightly linked 5′UTR SNP rs72696119 (delATTG linked with G), eQTL p values, and p values for pairwise t tests for differential expression by genotype. Expression for B (no significant difference) and T cells (individual level data not available) not shown.

individuals in the exposure (blood cell traits) dataset, we carried out a two-sample MR analysis. Two-sample MR supported a causal effect between blood cell counts and antibody response levels, consistent with direction of effect of the rs28362491:delATTG variant in the corresponding GWAS data. Increased RBC, basophil, neutrophil, monocyte, and platelet counts all associated with higher antibody responses, and increased lymphocyte and eosinophil counts associated with lower antibody responses (Figure 4; Table S14). Statistical power for two-sample MR for blood

cell traits and antibody responses was between 99% and 100% (Table S15), suggesting we had adequate power for these traits.

We did not identify any other significant associations between normalized blood cell counts and infection or inflammation-related disease (Table S16). However, we did not have sufficient power to deter a causal effect between these traits (2.5%–3%, Table S17). F statistics for the full cohort ranged from 103.028 for normalized lymphocyte count to 10.414 for platelet counts (Table S18).

Discussion

The genetic architecture underlying susceptibility to a host of infectious diseases is being resolved to greater detail. Loci, including the MHC, immunoglobulin loci, and *FUT2*, are becoming associated with a larger number of infectious traits when detected through multiple study formats, either through case-control analyses of clinical disease definition, self-report, or antigen-specific diagnostics.[15–17,19,23] Here, we identify rs28362491 in the promoter region of *NFKB1*, which is common in all populations (global allele frequency[del] = 0.42), that incurs a subtle, yet demonstrable effect across both the UK Biobank population and, where available, other population datasets, modulating the risk of a range of infections.

As expected, given the known pleiotropy of NF-κB signaling, our study provides significant evidence of the delicate balance between immune response to infection and inflammatory processes, where disruptions can tip the balance toward exacerbated immune responses and allergy or toward increased risk of infection and impaired resolution of inflammation. There are likely multiple mechanisms underlying this effect, involving differential hematopoiesis and immunoglobulin class switching driven by modulation of *NFKB1* gene expression.

On the basis of our functional work, we suggest that reduction of *NFKB1* expression, driven by the deletion variant, favors development of lymphoid progenitor cells and lymphoid cell lineages over myeloid lineages during hematopoiesis. NF-κB activity regulates B cell development and maturation and is important for T cell co-stimulation.[65] Despite moderate increases in the relative number of circulating lymphocytes, reduced *NFKB1* expression in T cells may result in impaired T cell function, thereby making them less responsive to infection. Variants tightly linked to rs28362491 have been associated with a reduced IgA/IgG:IgM ratio (Table S9), suggesting that reduced *NFKB1* expression may also affect B cell class switching during activation,[13,22] leading to the reduced IgG antibody levels observed in our study.

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

The observed inverse effect in allergy compared with infection and inflammatory disease may be a result of impaired resolution of inflammation stemming from a dysregulation of *NFKB1* expression as a result of disrupted promoter binding. NF-κB family members and downstream gene products play a key role in induction and regulation of inflammatory responses, and as such, are tightly regulated, possibly by relative ratios of the NF-κB p50-p65 (*NFKB1* encodes p50, p65 is encoded by *RELA*) and p50-p50 dimers.[54] Taken together with our cellular findings, it would appear that in the absence of haploinsufficiency, variation and regulation of expression is more important than overall *NFKB1* activity in altering risk of infection or inflammatory disease. However, follow-up studies will be required to fully dissect the effects of mild disruptions to NF-κB pathways, particularly since many of our observations are based on cross-sectional analyses, and therefore, longitudinal analyses should be prioritized for the future.

Although we present evidence from a variety of sources and study formats in support of our observation that strengthens the credibility of the association, our work does not prove definitively that, firstly, rs28362491:delATTG is the primary driving variant, nor that it is causal in all phenotypes studied. However, our fine-mapping methods would make it the most-likely candidate, and historic transcription binding studies would make it the most plausible functional modulator of expression. Causal inference through Mendelian randomization supports a possible link between hematopoietic pathways, cell function, and antibody responses, but the small effect sizes and variance explained by the rs28362491:delATTG variant mean that all but the most robust effects are likely to be missed in such analysis. Furthermore, these analyses rely on the assumption that the traits selected, and directionality of causation are correctly identified. Given the complexity of the NF-κB pathways and the complex nature of immune system interactions, it is possible that this assumption may not hold for all tested causal paths or all subsets of biobank participants at the time of data collection. As such, deeper validation of our hypothesis of mechanism will be highly challenging to dissect.

Our ability to dissect the relationship between antibody titer and pathogen-associated disease is limited by the small size of the current data subset and the low number of confirmed infections within the UK Biobank samples. While the multiplex serology platform allows us to capture

a measure of cumulative pathogen exposure, there are additional factors that will also affect the peak magnitude and speed of waning of antibody levels over time, including age, time since exposure or infection, chronic carriage of pathogens, and reactivation of latent infections. Longitudinal studies will provide better insights into how antibody titers change over time and with subsequent infection, which UK Biobank does offer opportunities to inspect, owing to follow up samples being available. We also recognize that the UK Biobank cohort is predominantly of European descent, and therefore not necessarily representative of the wider UK population.[66] As such, while we are able to draw correlations between antibody titers and reported disease, future studies should draw from a much wider cohort including participants of non-European descent, and individuals with known history of exposure.

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

Web resources

Supplemental information

Acknowledgments

Author contributions

Declaration of interests

The authors declare no competing interests.

References

1. Davila, S., Wright, V.J., Khor, C.C., Sim, K.S., Binder, A., Breunis, W.B., Inwald, D., Nadel, S., Betts, H., Carrol, E.D., et al. (2010). Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease. Nat. Genet. *42*, 772–776.

2. Gilchrist, J.J., Rautanen, A., Fairfax, B.P., Mills, T.C., Naranbhai, V., Trochet, H., Pirinen, M., Muthumbi, E., Mwarumba, S., Njuguna, P., et al. (2018). Risk of nontyphoidal Salmonella bacteraemia in African children is modified by STAT4. Nat. Commun. *9*, 1014.

3. Initiative, C.-H.G. (2021). Mapping the human genetic architecture of COVID-19. Nature *2021*, 1–8.

4. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., et al. (2020). Genetic mechanisms of critical illness in COVID-19. Nature *591*, 92–98.

5. The Severe Covid, G.G. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N. Engl. J. Med. *383*, 1522–1534.

6. Vergara, C., Thio, C.L., Johnson, E., Kral, A.H., O'Brien, T.R., Goedert, J.J., Mangia, A., Piazzolla, V., Mehta, S.H., Kirk,

G.D., et al. (2019). Multi-Ancestry Genome-Wide Association Study of Spontaneous Clearance of Hepatitis C Virus. Gastroenterology *156*, 1496–1507.e7.

7. Wang, Z., Sun, Y., Fu, X., Yu, G., Wang, C., Bao, F., Yue, Z., Li, J., Sun, L., Irwanto, A., et al. (2016). A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. Nat. Commun. *7*, 13760.

8. Zhang, F., Liu, H., Chen, S., Low, H., Sun, L., Cui, Y., Chu, T., Li, Y., Fu, X., Yu, Y., et al. (2011). Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. Nat. Genet. *43*, 1247–1251.

9. Duggal, P., Thio, C.L., Wojcik, G.L., Goedert, J.J., Mangia, A., Latanich, R., Kim, A.Y., Lauer, G.M., Chung, R.T., Peters, M.G., et al. (2013). Genome-Wide Association Study of Spontaneous Resolution of Hepatitis C Virus Infection: Data From Multiple Cohorts. Ann. Intern. Med. *158*, 235–245.

10. Dunstan, S.J., Hue, N.T., Han, B., Li, Z., Tram, T.T.B., Sim, K.S., Parry, C.M., Chinh, N.T., Vinh, H., Lan, N.P.H., et al. (2014). Variation at HLA-DRB1 is associated with resistance to enteric fever. Nat. Genet. *46*, 1333–1336.

11. Tian, C., Hromatka, B.S., Kiefer, A.K., Eriksson, N., Noble, S.M., Tung, J.Y., and Hinds, D.A. (2017). Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat. Commun. *8*, 599.

12. VanBlargan, L.A., Goo, L., and Pierson, T.C. (2016). Deconstructing the Antiviral Neutralizing-Antibody Response: Implications for Vaccine Development and Immunity. Microbiol. Mol. Biol. Rev. *80*, 989–1010.

13. Jonsson, S., Sveinbjornsson, G., de Lapuente Portilla, A.L., Swaminathan, B., Plomp, R., Dekkers, G., Ajore, R., Ali, M., Bentlage, A.E.H., Elmér, E., et al. (2017). Identification of sequence variants influencing immunoglobulin levels. Nat. Genet. *49*, 1182–1191.

14. Png, E., Thalamuthu, A., Ong, R.T.H., Snippe, H., Boland, G.J., and Seielstad, M. (2011). A genome-wide association study of hepatitis B vaccine response in an Indonesian population reveals multiple independent risk variants in the HLA region. Hum. Mol. Genet. *20*, 3893–3898.

15. Hammer, C., Begemann, M., McLaren, P.J., Bartha, I., Michel, A., Klose, B., Schmitt, C., Waterboer, T., Pawlita, M., Schulz, T.F., et al. (2015). Amino Acid Variation in HLA Class II Proteins Is a Major Determinant of Humoral Response to Common Viruses. Am. J. Hum. Genet. *97*, 738–743.

16. Hodel, F., Chong, A.Y., Scepanovic, P., Xu, Z.M., Naret, O., Thorball, C.W., Rüeger, S., Marques-Vidal, P., Vollenweider, P., Begemann, M., et al. (2021). Human genomics of the humoral immune response against polyomaviruses. Virus Evol. *7*, veab058.

17. Kachuri, L., Francis, S.S., Morrison, M.L., Wendt, G.A., Bossé, Y., Cavazos, T.B., Rashkin, S.R., Ziv, E., and Witte, J.S. (2020). The landscape of host genetic factors involved in immune response to common viral infections. Genome Med. *12*, 93.

18. Andreu-Sánchez, S., Bourgonje, A.R., Vogl, T., Kurilshikov, A., Leviatan, S., Ruiz-Moreno, A.J., Hu, S., Sinha, T., Vich Vila, A., Klompus, S., et al. (2023). Phage display sequencing reveals that genetic, environmental, and intrinsic factors influence variation of human antibody epitope repertoire. Immunity *56*, 1376–1392.e1378.

19. Venkataraman, T., Valencia, C., Mangino, M., Morgenlander, W., Clipman, S.J., Liechti, T., Valencia, A., Christofidou, P., Spector, T., Roederer, M., et al. (2022). Analysis of antibody binding specificities in twin and SNP-genotyped cohorts reveals that antiviral antibody epitope selection is a heritable trait. Immunity *55*, 174–184.e5.

20. Hayden, M.S., and Ghosh, S. (2011). NF-kappaB in immunobiology. Cell Res. *21*, 223–244.

21. Hayden, M.S., West, A.P., and Ghosh, S. (2006). NF-kappaB and the immune response. Oncogene *25*, 6758–6780.

22. Tuijnenburg, P., Lango Allen, H., Burns, S.O., Greene, D., Jansen, M.H., Staples, E., Stephens, J., Carss, K.J., Biasci, D., Baxendale, H., et al. (2018). Loss-of-function nuclear factor kappaB subunit 1 (NFKB1) variants are the most common monogenic cause of common variable immunodeficiency in Europeans. J. Allergy Clin. Immunol. *142*, 1285–1296.

23. Butler-Laporte, G., Kreuzer, D., Nakanishi, T., Harroud, A., Forgetta, V., and Richards, J.B. (2020). Genetic Determinants of Antibody-Mediated Immune Responses to Infectious Diseases Agents: A Genome-Wide and HLA Association Study. Open Forum Infect. Dis. *7*, ofaa450.

24. Mentzer, A.J., Brenner, N., Allen, N., Littlejohns, T.J., Chong, A.Y., Cortes, A., Almond, R., Hill, M., Sheard, S., McVean, G., et al. (2022). Identification of host–pathogen-disease relationships using a scalable multiplex serology platform in UK Biobank. Nat. Commun. *13*, 1818.

25. Waterboer, T., Sehr, P., Michael, K.M., Franceschi, S., Nieland, J.D., Joos, T.O., Templin, M.F., and Pawlita, M. (2005). Multiplex Human Papillomavirus Serology Based on In Situ–Purified Glutathione S-Transferase Fusion Proteins. Clin. Chem. *51*, 1845–1853.

26. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.

27. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

28. Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J.L., Danecek, P., Malerba, G., Trabetti, E., Zheng, H.F., et al. (2015). Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat. Commun. *6*, 8111.

29. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. *48*, 1443–1448.

30. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.

31. Loh, P.R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet. *47*, 284–290.

32. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience *4*, 7.

33. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.

34. Benner, C., Spencer, C.C.A., Havulinna, A.S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics *32*, 1493–1501.

35. The Wellcome Trust Case Control, C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.

36. Firmann, M., Mayor, V., Vidal, P.M., Bochud, M., Pécoud, A., Hayoz, D., Paccaud, F., Preisig, M., Song, K.S., Yuan, X., et al. (2008). The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. BMC Cardiovasc. Disord. *8*, 6.

37. Han, B., and Eskin, E. (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am. J. Hum. Genet. *88*, 586–598.

38. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat. Genet. *50*, 1335–1341.

39. Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., and Knight, J.C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science *343*, 1246949.

40. Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O., and Knight, J.C. (2012). Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet. *44*, 502–510.

41. Gilchrist, J.J., Makino, S., Naranbhai, V., Sharma, P.K., Koturan, S., Tong, O., Taylor, C.A., Watson, R.A., de los Aires, A.V., Cooper, R., et al. (2022). Natural Killer cells demonstrate distinct eQTL and transcriptome-wide disease associations, highlighting their role in autoimmunity. Nat. Commun. *13*, 4073.

42. Naranbhai, V., Fairfax, B.P., Makino, S., Humburg, P., Wong, D., Ng, E., Hill, A.V.S., and Knight, J.C. (2015). Genomic modulators of gene expression in human neutrophils. Nat. Commun. *6*, 7545.

43. Schmiedel, B.J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A.G., White, B.M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J.A., McVicker, G., et al. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell *175*, 1701–1715.e16.

44. Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet. *17*, e1009440.

45. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. J. R. Stat. Soc. Series B Stat. Methodol. *82*, 1273–1300.

46. Zou, Y., Carbonetto, P., Wang, G., and Stephens, M. (2021). Fine-mapping from summary data with the "Sum of Single Effects" model. Preprint at bioRxiv.

47. Burgess, S., Davies, N.M., and Thompson, S.G. (2016). Bias due to participant overlap in two-sample Mendelian randomization. Genet. Epidemiol. *40*, 597–608.

48. Broadbent, J.R., Foley, C.N., Grant, A.J., Mason, A.M., Staley, J.R., and Burgess, S. (2020). MendelianRandomization v0.5.0: updates to an R package for performing Mendelian randomization analyses using summarized data. Wellcome Open Res. *5*, 252.

49. Yavorska, O.O., and Burgess, S. (2017). MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int. J. Epidemiol. *46*, 1734–1739.

50. Scepanovic, P., Alanio, C., Hammer, C., Hodel, F., Bergstedt, J., Patin, E., Thorball, C.W., Chaturvedi, N., Charbit, B., Abel, L., et al. (2018). Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. Genome Med. *10*, 59.

51. Karban, A.S., Okazaki, T., Panhuysen, C.I.M., Gallegos, T., Potter, J.J., Bailey-Wilson, J.E., Silverberg, M.S., Duerr, R.H., Cho, J.H., Gregersen, P.K., et al. (2004). Functional annotation of a novel NFKB1 promoter polymorphism that increases risk for ulcerative colitis. Hum. Mol. Genet. *13*, 35–45.

52. Jenner, R.G., and Young, R.A. (2005). Insights into host responses against pathogens from transcriptional profiling. Nat. Rev. Microbiol. *3*, 281–294.

53. Seidi, A., Mirzaahmadi, S., Mahmoodi, K., and Soleiman-Soltanpour, M. (2018). The association between NFKB1 -94ATTG ins/del and NFKB1A 826C/T genetic variations and coronary artery disease risk. Mol. Biol. Res. Commun. *7*, 17–24.

54. Mishra, A., Srivastava, A., Mittal, T., Garg, N., and Mittal, B. (2013). Role of inflammatory gene polymorphisms in left ventricular dysfunction (LVD) susceptibility in coronary artery disease (CAD) patients. Cytokine *61*, 856–861.

55. Lai, H.-M., Li, X.-M., Yang, Y.-N., Ma, Y.-T., Xu, R., Pan, S., Zhai, H., Liu, F., Chen, B.-D., and Zhao, Q. (2015). Genetic Variation in NFKB1 and NFKBIA and Susceptibility to Coronary Artery Disease in a Chinese Uygur Population. PLoS One *10*, e0129144.

56. Bajwa, E.K., Cremer, P.C., Gong, M.N., Zhai, R., Su, L., Thompson, B.T., and Christiani, D.C. (2011). An NFKB1 promoter insertion/deletion polymorphism influences risk and outcome in acute respiratory distress syndrome among Caucasians. PLoS One *6*, e19469.

57. Borm, M.E.A., van Bodegraven, A.A., Mulder, C.J.J., Kraal, G., and Bouma, G. (2005). A NFKB1 promoter polymorphism is involved in susceptibility to ulcerative colitis. Int. J. Immunogenet. *32*, 401–405.

58. Ellinghaus, D., Jostins, L., Spain, S.L., Cortes, A., Bethune, J., Han, B., Park, Y.R., Raychaudhuri, S., Pouget, J.G., Hübenthal, M., et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat. Genet. *48*, 510–518.

59. Kawashima, M., Hitomi, Y., Aiba, Y., Nishida, N., Kojima, K., Kawai, Y., Nakamura, H., Tanaka, A., Zeniya, M., Hashimoto, E., et al. (2017). Genome-wide association studies identify PRKCB as a novel genetic susceptibility locus for primary biliary cholangitis in the Japanese population. Hum. Mol. Genet. *26*, 650–659.

60. Dudding, T., Haworth, S., Lind, P.A., Sathirapongsasuti, J.F., 23andMe Research Team, Tung, J.Y., Mitchell, R., Colodro-Conde, L., Medland, S.E., Gordon, S., et al. (2019). Genome wide analysis for mouth ulcers identifies associations at immune regulatory loci. Nat. Commun. *10*, 1052.

61. Waage, J., Standl, M., Curtin, J.A., Jessen, L.E., Thorsen, J., Tian, C., Schoettler, N., 23andMe Research Team; and AAGC collaborators, and Flores, C., et al. (2018). Genome-wide association and HLA fine-mapping studies identify risk loci and

genetic pathways underlying allergic rhinitis. Nat. Genet. *50*, 1072–1080.

62. Johansson, Å., Rask-Andersen, M., Karlsson, T., and Ek, W.E. (2019). Genome-wide association analysis of 350 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema. Hum. Mol. Genet. *28*, 4022–4041.

63. Wang, D., Xie, T., Xu, J., Wang, H., Zeng, W., Rao, S., Zhou, K., Pei, F., and Zhou, Z. (2016). Genetic association between NFKB1 -94 ins/del ATTG Promoter Polymorphism and cancer risk: a meta-analysis of 42 case-control studies. Sci. Rep. *6*, 30220.

64. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.

65. Kaileh, M., and Sen, R. (2012). NF-kappaB function in B lymphocytes. Immunol. Rev. *246*, 254–271.

66. Keyes, K.M., and Westreich, D. (2019). UK Biobank, big data, and the consequences of non-representativeness. Lancet *393*, 1297.