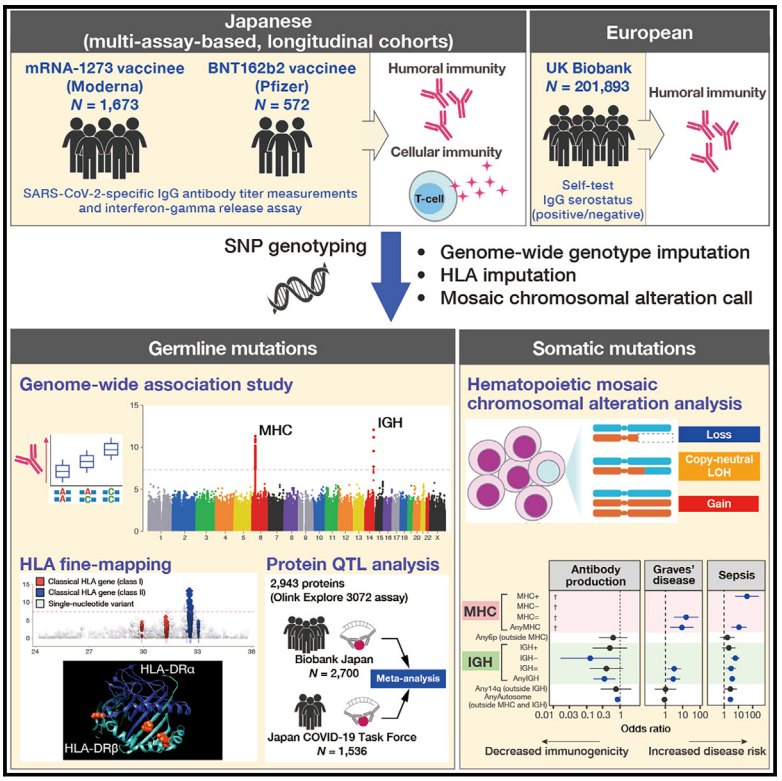
Cell Genomics

Article

Germline variants and mosaic chromosomal alterations affect COVID-19 vaccine immunogenicity

Graphical abstract



Highlights

- GWAS of vaccine immunogenicity reveals the contribution of MHC and IGH loci
- Protein QTL analysis identifies circulating immune regulators modulated by these loci
- Hematopoietic somatic alterations affecting MHC/IGH impair vaccine immunogenicity
- These hematopoietic somatic alterations confer infectious/immune disease risk

Authors

Kyuto Sonehara, Yoshifumi Uwamino, Ryunosuke Saiki, ..., Seishi Ogawa, Yukinori Okada, Ho Namkoong

Correspondence

uwamino@keio.jp (Y.U.),
yuki-okada@m.u-tokyo.ac.jp (Y.O.),
hounamugun@keio.jp (H.N.)

In brief

Sonehara and Uwamino et al. perform a genome-wide association study of COVID-19 vaccine immunogenicity, finding associations at the MHC and IGH loci. Hematopoietic mosaic chromosomal alterations affecting these loci impair vaccine immunogenicity and increase infectious/immune disease risk. This revealed the contribution of both germline and somatic mutations to adaptive immunity.

Sonehara et al., 2025, Cell Genomics 5, 100783
March 12, 2025
https://doi.org/10.1016/j.xgen.2025.100783

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Germline variants and mosaic chromosomal alterations affect COVID-19 vaccine immunogenicity

Kyuto Sonehara,[1,2,3,44,45] Yoshifumi Uwamino,[4,45,46,*] Ryunosuke Saiki,[5] Masaru Takeshita,[6] Shinichi Namba,[1,2,3] Shunsuke Uno,[7] Tomoko Nakanishi,[1,8] Tomoyasu Nishimura,[7,9] Tatsuhiko Naito,[2,3] Go Sato,[2,3,10] Masahiro Kanai,[11,12,13,14] Aoxing Liu,[11,12,13,15,16] Sho Uchida,[7] Toshinobu Kurafuji,[17] Akiko Tanabe,[17] Tomoko Arai,[17] Akemi Ohno,[17] Ayako Shibata,[4] Shiho Tanaka,[4] Masatoshi Wakui,[4] Shoko Kashimura,[7] Chiharu Tomi,[7] Akemi Hara,[7]

(Author list continued on next page)

[1]Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[2]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan
[3]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
[4]Department of Laboratory Medicine, Keio University School of Medicine, Tokyo, Japan
[5]Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan
[6]Division of Rheumatology, Department of Internal Medicine, Keio University School of Medicine, Tokyo, Japan
[7]Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan
[8]Japan Society for the Promotion of Science, Tokyo, Japan
[9]Keio University Health Center, Shinjuku-ku, Tokyo, Japan
[10]Department of Gastroenterological Surgery, Osaka University Graduate School of Medicine, Suita, Japan
[11]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
[12]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[13]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland
[14]Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA
[15]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[16]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[17]Clinical Laboratory, Keio University Hospital, Tokyo, Japan
[18]Division of Pharmacodynamics, Keio University Faculty of Pharmacy, Tokyo, Japan
[19]Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo, Japan
[20]Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, Suita, Japan
[21]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan

(Affiliations continued on next page)

SUMMARY

Vaccine immunogenicity is influenced by the vaccinee's genetic background. Here, we perform a genome-wide association study of vaccine-induced SARS-CoV-2-specific immunoglobulin G (IgG) antibody titers and T cell immune responses in 1,559 mRNA-1273 and 537 BNT162b2 vaccinees of Japanese ancestry. SARS-CoV-2-specific antibody titers are associated with the immunoglobulin heavy chain (IGH) and major histocompatibility complex (MHC) locus, and T cell responses are associated with MHC. The lead variants at IGH contain a population-specific missense variant (rs1043109-C; p.Leu192Val) in the immunoglobulin heavy constant gamma 1 gene (*IGHG1*), with a strong decreasing effect ($\beta = -0.54$). Antibody-titer-associated variants modulate circulating immune regulatory proteins (e.g., LILRB4 and FCRL6). Age-related hematopoietic expanded mosaic chromosomal alterations (mCAs) affecting MHC and IGH also impair antibody production. MHC-/IGH-affecting mCAs confer infectious and immune disease risk, including sepsis and Graves' disease. Impacts of expanded mosaic loss of chromosomes X/Y on these phenotypes were examined. Altogether, both germline and somatic mutations contribute to adaptive immunity functions.

INTRODUCTION

Vaccination has played a pivotal role in reducing the morbidity of and mortality from infectious diseases, significantly improving public health worldwide, as demonstrated during the recent COVID-19 pandemic. Vaccine efficacy is influenced by the factors of the vaccinees.[1] Understanding these factors is crucial for preparing for the next pandemic, as it will aid in developing

Shiori Yoshikawa,[7] Keiko Gotanda,[7] Kana Misawa,[7,18] Hiromu Tanaka,[19] Shuhei Azekawa,[19] Qingbo S. Wang,[1,2] Ryuya Edahiro,[2,3,20] Yuya Shirai,[2,20,21] Kenichi Yamamoto,[2,22,23] Genta Nagao,[1,19] Takuo Suzuki,[24] Masato Kiyoshi,[24] Akiko Ishii-Watabe,[24] Shinichi Higashiue,[25] Shuzo Kobayashi,[25] Hiroki Yamaguchi,[26] Yasushi Okazaki,[27] Naoyuki Matsumoto,[27] Akihide Masumoto,[28] Hidenobu Koga,[28] Akinori Kanai,[29] Japan COVID-19 Task Force, Biobank Japan Project, Yoshiya Oda,[30] Yutaka Suzuki,[29] Koichi Matsuda,[31] Yuko Kitagawa,[32] Ryuji Koike,[33] Akinori Kimura,[34] Atsushi Kumanogoh,[20,35] Akihiko Yoshimura,[36] Seiya Imoto,[37] Satoru Miyano,[38] Takanori Kanai,[39] Koichi Fukunaga,[19] Naoki Hasegawa,[7] Mitsuru Murata,[4,40] Hiromichi Matsushita,[4] Seishi Ogawa,[5,41,42] Yukinori Okada,[1,2,3,21,43,46,47,*] and Ho Namkoong[7,46,*]

[22]Department of Pediatrics, Osaka University Graduate School of Medicine, Suita, Japan
[23]Division of Health Science, Osaka University Graduate School of Medicine, Suita, Japan
[24]Division of Biological Chemistry and Biologicals, National Institute of Health Sciences, Kanagawa, Japan
[25]Tokushukai Group, Tokyo, Japan
[26]Department of Hematology, Nippon Medical School, Tokyo, Japan
[27]Diagnostics and Therapeutics of Intractable Diseases, Intractable Disease Research Center, Graduate School of Medicine, Juntendo University, Tokyo, Japan
[28]ASO Iizuka Hospital, Iizuka, Fukuoka, Japan
[29]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan
[30]Department of Lipidomics, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan
[31]Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
[32]Department of Surgery, Keio University School of Medicine, Tokyo, Japan
[33]Health Science Research and Development Center (HeRD), Tokyo Medical and Dental University, Tokyo, Japan
[34]Institute of Research, Tokyo Medical and Dental University, Tokyo, Japan
[35]Department of Immunopathology, Immunology Frontier Research Center, Osaka University, Suita, Japan
[36]Department of Microbiology and Immunology, Keio University School of Medicine, Tokyo, Japan
[37]Division of Health Medical Intelligence, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
[38]M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan
[39]Division of Gastroenterology and Hepatology, Department of Medicine, Keio University School of Medicine, Tokyo, Japan
[40]Research Center of Clinical Medicine, International University of Health and Welfare, Tokyo, Japan
[41]Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan
[42]Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden
[43]Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Japan
[44]Present address: Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK
[45]These authors contributed equally
[46]These authors contributed equally
[47]Lead contact
*Correspondence: uwamino@keio.jp (Y.U.), yuki-okada@m.u-tokyo.ac.jp (Y.O.), hounamugun@keio.jp (H.N.)
https://doi.org/10.1016/j.xgen.2025.100783

effective vaccines and devising efficient vaccination strategies. During the COVID-19 pandemic, the COVID-19 vaccines were administered on an unprecedented scale to the general population, not limited to specific attributes such as age group, providing a unique opportunity to comprehensively study the relationship between vaccinees' factors and vaccine immunogenicity.

Beyond age and sex, studies on host genetic determinants of vaccine antibody production have identified the major histocompatibility complex (MHC) locus as a significant factor in European populations.[2,3] Given that infectious pathogens have been a major selective pressure in human evolutionary history, shaping the population-specific allele frequency spectrum in modern populations, it is highly likely that population-specific genetic variants contribute to the immune response to novel immunogens.[4,5] Indeed, international host genetic studies of SARS-CoV-2 infection have highlighted population-specific variants.[6,7] Similarly, COVID-19 vaccines represent a new and massive exposure factor for humanity, suggesting a high likelihood of detecting population-specific variants in response to vaccination. Despite this, genetic studies of the immune response to vaccines in cohorts of non-European ancestry have been insufficient.

In addition to germline mutations, somatic mutations, including hematopoietic mosaic chromosomal alterations (mCAs), have recently gained attention for their role in influencing phenotypic traits. Traditionally, age-related somatic mutations have been studied primarily in the context of malignant tumors. However, recent research has highlighted their potential involvement in inflammatory and non-malignant diseases, such as severe COVID-19 and heart failure.[8–10] Furthermore, there is a growing body of research emphasizing the importance of analyzing both germline and somatic mutations together.[11] These reports have led us to hypothesize that, in addition to germline mutations, age-related somatic mutations may also significantly affect the immune response to COVID-19 vaccines.

Here, we established multiassay-based prospective vaccination cohorts of Japanese ancestry, including recipients of the Pfizer and Moderna COVID-19 mRNA vaccines (Figure 1). Evaluating immune responses following vaccination inherently involves several challenges, one of which is the variability in

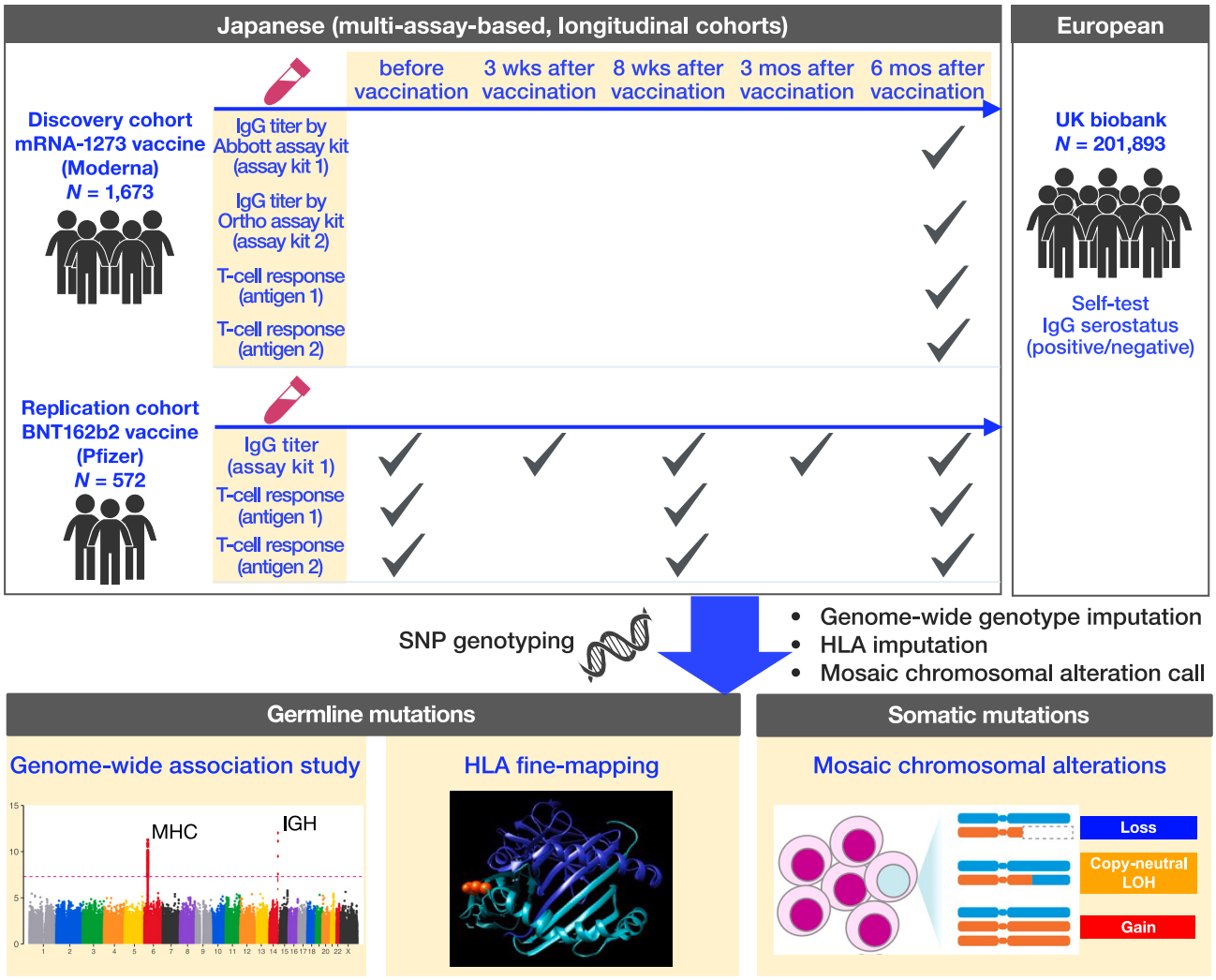


Figure 1. Overview of the study

We enrolled 1,673 individuals receiving the mRNA-1273 COVID-19 mRNA vaccine (Moderna) and 572 individuals receiving the BNT162b2 COVID-19 mRNA vaccine (Pfizer), both of Japanese ancestry. By measuring COVID-19 vaccine-induced IgG antibody titers (humoral immunity) and T cell response levels (cellular immunity) using multiple assay kits at multiple time points, we performed genome-wide association studies (GWASs) and HLA fine-mapping to identify germline mutations affecting vaccine immunogenicity. We compared associated loci in the Japanese cohorts with those in the UK Biobank dataset. Finally, we demonstrated that hematopoietic mosaic chromosomal alterations affecting the genetic loci implicated in GWASs contribute to aging-related vaccine efficacy impairment.

antibody titer measurements across different diagnostic kits. To address this, we evaluated the immune response using multiple kits to ensure evaluation consistency.[12] In addition, beyond assessing humoral immunity through antibody titers, quantifying the cellular immune response is another challenge. To tackle this, we conducted a large-scale analysis using the interferon-gamma release assay to quantify SARS-CoV-2-specific T cell immune responses, which measures cellular immunity.[13]

Leveraging this extensive resource, we conducted a series of genome-wide association studies (GWASs), reproducibly demonstrating the contribution of genetic variants at the immunoglobulin heavy chain (IGH) locus as well as MHC to vaccine immunogenicity. These loci were found to be effective not only in the Japanese population but also among Europeans, as corroborated by data from the UK Biobank (UKB). Moreover, the impact of these genetic loci extends beyond COVID-19 vaccines, influencing responses to influenza vaccines and indicating its significance across different populations and vaccine types. In addition to germline mutations, we discovered that hematopoietic mCA events affecting MHC and IGH influence vaccine efficacy. Furthermore, these mCA events conferred risk of infectious and immune diseases, suggesting the contribution of both germline and somatic mutations in these loci to the general human immune response and the pathogenesis and prognosis of immune and infectious diseases. We further evaluated the influence of mosaic loss of chromosome X in females (mLOX) and

Table 1. Association summary of the lead variants in the Japanese population

Trait	Locus	ID	Chr	Position (GRCh37)	Alleles	Effect allele	Discovery cohort				Replication cohort				Meta-analysis			
							EAF	Effect size	SE	*p*	EAF	Effect size	SE	*p*	EAF	Effect size	SE	*p*
IgG production	MHC	rs9270585	6	32,561,300	C/T	T	0.53	0.19	0.032	2.7×10^{-9}	0.53	0.15	0.052	0.0040	0.53	0.18	0.027	5.0×10^{-11}
	IGH	rs1043249	14	106,208,306	A/G	G	0.023	−0.66	0.10	2.2×10^{-10}	0.035	−0.42	0.14	0.0032	0.027	−0.57	0.084	6.9×10^{-12}
	IGH	rs1043109	14	106,208,326	G/C	C	0.023	−0.65	0.10	3.5×10^{-10}	0.034	−0.35	0.14	0.013	0.027	−0.54	0.083	6.8×10^{-11}
	IGH	rs193160354	14	106,208,327	G/A	A	0.023	−0.65	0.10	3.4×10^{-10}	0.034	−0.35	0.14	0.013	0.027	−0.54	0.083	6.8×10^{-11}
T cell response	MHC	rs9282156	6	32,633,918	C/CAA	CAA	0.74	−0.43	0.037	1.5×10^{-30}	0.73	−0.36	0.063	7.4×10^{-9}	0.74	−0.41	0.032	1.1×10^{-37}

EAF, effect allele frequency.
The three lead variants at IGH are shown. rs1043109 is a missense SNP (p.Leu192Val) in *IGHG1*, while rs1043249 and rs193160354 are synonymous SNPs.

Y in males (mLOY), the more common and sex-specific clonal somatic alterations, on these phenotypes.

Understanding the factors within populations that determine vaccine antibody titers is crucial as the world continues to develop various vaccines in response to pandemics. Herein, we report new perspectives on the germline and somatic underpinnings of vaccine efficacy and immune-related diseases.

RESULTS

GWAS of COVID-19 vaccine immunogenicity

We recruited 1,673 individuals as a discovery cohort and 572 individuals as a replication cohort, both of Japanese ancestry (Figures 1 and S1). The individuals in the discovery cohort received two doses of the mRNA-1273 COVID-19 mRNA vaccine (Spikevax intramuscular injection, Moderna, Cambridge, MA, USA) between June and July 2021, and those in the replication cohort received two doses of the BNT162b2 COVID-19 mRNA vaccine (COMIRNATY intramuscular injection, Pfizer, New York, NY, USA) between February and April 2021. The demographic characteristics of the two vaccination cohorts are shown in Table S1. In both cohorts, we collected blood samples 6 months after the vaccination to evaluate variations in the vaccine immunogenicity. We quantified vaccine immunogenicity separately as immunoglobulin G (IgG) antibody titers against the SARS-CoV-2 spike protein (i.e., humoral immunity) and SARS-CoV-2 antigen-specific T cell response levels (i.e., cellular immunity). To ensure the technical robustness of the vaccine immunogenicity measurements and subsequent genetic association analyses, our study design adopted multiple independent assay kits (Figure S2). Specifically, IgG antibody titers were measured by two different assay kits: the Abbott SARS-CoV-2 IgG II quantitative antibody assay (assay kit 1) and the Ortho VITROS SARS-CoV-2 spike-specific quantitative IgG assay (assay kit 2) (STAR Methods). T cell responses were measured using the QuantiFERON SARS-CoV-2 assay with two different antigens derived from SARS-CoV-2 (antigen 1 and antigen 2). DNA from all the study participants was genotyped using the Infinium Asian screening array, which yielded 537,027 directly genotyped SNPs after stringent quality control (QC). After genome-wide genotype imputation using a population-specific reference panel (n = 11,754 Japanese; STAR Methods), we obtained 8,956,255 high-quality genetic variants available for GWAS (imputation *INFO* > 0.7 and minor allele frequency [MAF] > 0.5%).

To investigate the genetic architecture of the vaccine-induced humoral immunity, we performed an association analysis on the SARS-CoV-2-specific IgG antibody titers measured using assay kit 1 in the discovery cohort (n = 1,559). We identified genome-wide significant associations at the MHC locus at 6p21 ($p = 2.7 \times 10^{-9}$, $\beta = 0.19$, and SE = 0.032 at rs9270585 [6:32561300:C:T]) and the IGH locus at 14q32 ($p = 2.2 \times 10^{-10}$, $\beta = -0.66$, and SE = 0.10 at rs1043249 [14:106208306:A:G]). As a replication analysis, we performed an association analysis in the independent replication cohort (n = 537), confirming significant associations of both loci with concordant directional effects (p = 0.0040 for rs9270585 and p = 0.0032 for rs1043249; Table 1; meta-analysis results shown in Figure 2A; regional plots shown in Figure S3). Given that the

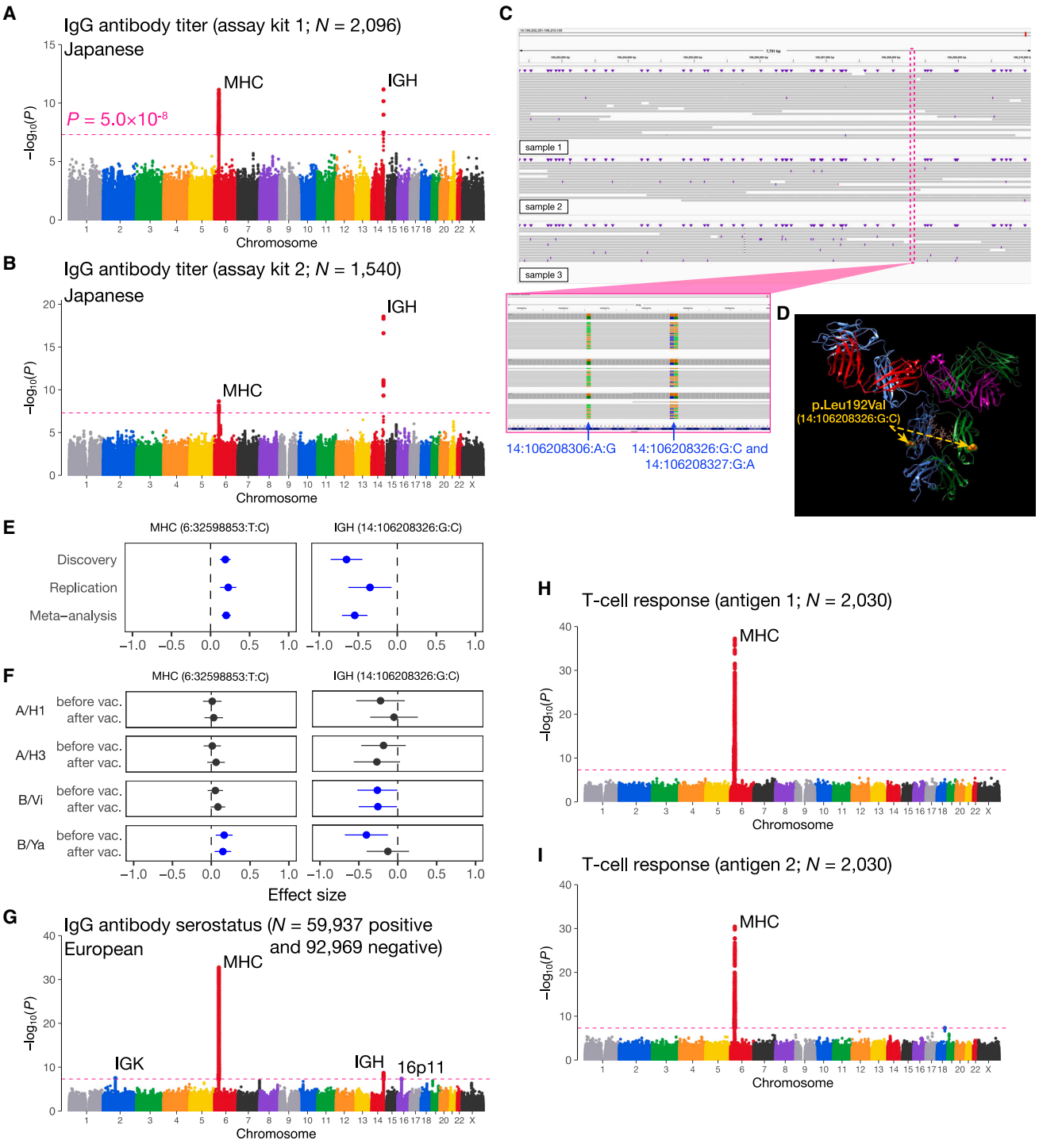


Figure 2. Genome-wide association study of COVID-19 vaccine immunogenicity

(A) Manhattan plot for the GWAS meta-analysis of the IgG antibody titer measured by assay kit 1.

(B) Manhattan plot for the GWAS of the IgG antibody titer measured by assay kit 2.

(C) The lead variants in *IGHG1* were validated by PacBio HiFi long-read sequencing on the three samples selected from the imputation reference panel. A 64 bp window around the three lead variants (14:106208306:A:G, 14:106208326:G:C, and 14:106208327:G:A) is highlighted as the pink dashed box and enlarged.

(D) *IGHG1* amino acid position p.Leu192Val is highlighted as orange-colored spheres. The protein structure of the human IgG is based on Protein Data Bank entry PDB: 1HZH and was prepared using UCSF Chimera (version 1.16).[14]

(E) Forest plots of the lead variant associations with the IgG antibody titer measured by assay kit 1. Blue markers indicate $p < 0.05$.

(legend continued on next page)

above analyses were both based on the SARS-CoV-2-specific IgG antibody titer measurements using assay kit 1, to rule out potential assay-specific effects and confirm the robustness of our findings, we further performed a genome-wide association analysis on the SARS-CoV-2-specific IgG antibody titers measured by assay kit 2 in the discovery cohort. In this technical validation analysis (n = 1,540), we confirmed significant associations at the same loci at MHC and IGH ($p = 1.0 \times 10^{-8}$ for rs9270585 and $p = 2.9 \times 10^{-19}$ for rs1043249; Figure 2B). The lead variant at MHC in the assay kit 1 GWAS (rs9270585) was in partial linkage disequilibrium (LD) with the lead variant in the assay kit 2 GWAS (rs9270613 [6:32565109:T:C]; $r^2 = 0.42$ in the Japanese imputation reference panel). Nevertheless, the association statistics within MHC showed a strong correlation between assay kits 1 and 2 (Pearson $r = 0.97$), suggesting that the difference in the lead variants is due to statistical fluctuations (Figure S4). The lead variant at IGH in the assay kit 1 GWAS (rs1043249) was also found as the lead variant in the assay kit 2 GWAS.

The IGH locus is one of the most structurally polymorphic and complex regions in the human genome, making it potentially challenging to accurately call the genetic variants using conventional short-read whole-genome sequencing (WGS), by which the GWAS imputation reference panel was constructed. To validate the short-read WGS-based calls, we selected three samples heterozygous for the lead variants at IGH from the imputation reference panel and performed PacBio HiFi long-read WGS on them. We confirmed that there were no structural variants suspected to cause spurious variant calls around the variants and that the variants were consistently genotyped using this orthogonal sequencing technology (Figure 2C). We also note that the GWAS individuals were all directly genotyped using Sanger sequencing for the three lead variants at IGH (STAR Methods). Among the three lead variants at IGH that are in nearly perfect LD ($r^2 \geq 0.97$ for any variant pair), rs1043109 (14:106208326:G:C) is a missense SNP (p.Leu192Val) in the immunoglobulin heavy constant gamma 1 gene (*IGHG1*), while the other two variants are coding but synonymous SNPs. *IGHG1* encodes the constant region of IgG1, the major subclass of IgG. The amino acid substitution p.Leu192Val is located outside of the IgG1 CH2 domain, the contact site for the neonatal Fc receptor, suggesting its functional consequence in regulating IgG turnover (Figure 2D). The strong decreasing effect ($\beta = -0.54$ and SE = 0.083 on the standard normal distribution scale; STAR Methods) also suggested the causal role of this missense SNP. Notably, the antibody titer-decreasing allele, rs1043109-C, is exclusively prevalent in East Asian people according to gnomAD v.4.0.0[15] (0.026 in the East Asian group but <0.005 in the other groups; Figure S5), underscoring the importance of genetic studies performed in ancestrally diverse populations not limited to Europeans.

To ask whether our finding is generalizable to vaccines against viruses other than SARS-CoV-2, we also measured the antibody titers against influenza viruses in the replication cohort before and after influenza vaccination (STAR Methods). We note that the antibody titers were measured using hemagglutination inhibition assay, the standard assay to titrate influenza antibodies based on the agglutination of red blood cells mediated by various types of globulins not limited to IgG. We tested the associations of the MHC and IGH variants with the antibody titers against four influenza subtypes: influenza A virus subtype H1N1, influenza A virus subtype H3N2, influenza virus subtype B/Victoria, and influenza virus subtype B/Yamagata. In this analysis, we tested rs3104382 (6:32598853:T:C), the lead variant at MHC in the meta-analysis combining the discovery and replication cohorts, and rs1043109, the presumably functional missense SNP at IGH (Figure 2E). The MHC variant significantly increased the antibody titer against B/Yamagata ($p = 0.0027$, $\beta = 0.17$, and SE = 0.055 before vaccination and $p = 0.0059$, $\beta = 0.15$, and SE = 0.055 after vaccination), and the IGH variant significantly decreased the antibody titer against B/Victoria ($p = 0.046$, $\beta = -0.26$, and SE = 0.13 before vaccination and $p = 0.042$, $\beta = -0.26$, and SE = 0.13 after vaccination) and B/Yamagata ($p = 0.0043$, $\beta = -0.40$, and SE = 0.14 before vaccination). Given the strong and complex LD pattern in MHC, we further performed a human leukocyte antigen (HLA) fine-mapping analysis on the COVID-19 vaccine association and the influenza B/Yamagata association (STAR Methods). While the class II HLA gene variants were top associated in MHC with both the COVID-19 and the influenza antibody titers, the lead HLA variant was HLA-DQβ1 amino acid position 203 ($p_{omnibus} = 9.5 \times 10^{-8}$), different from that for COVID-19 IgG antibody (HLA-DRβ1 amino acid position 96, as discussed later) (Figure S6A). Even when we conditioned on HLA-DRβ1 amino acid position 96, HLA-DQβ1 amino acid position 203 remained the top-associated HLA variant ($p_{omnibus} = 2.5 \times 10^{-5}$; Figure S6B), suggesting that the associated HLA variants are different between COVID-19 and influenza.

Overall, the effect sizes were directionally concordant with those for COVID-19 vaccines, suggesting that the IGH variant may be a genetic determinant of acute IgG production in the immediate response to vaccination in general not limited to SARS-CoV-2-specific IgG antibody (Figure 2F). The variant effects for the influenza vaccine were present both before and after the vaccination. In contrast, those for COVID-19 vaccines were present only after the vaccination (as described later). This difference is presumably because most of the participants were exposed to influenza vaccines or viruses before this study period, unlike the newly developed COVID-19 vaccines.

To better understand the genetic architecture of vaccine-induced antibody production beyond ancestry, we referred to

(F) Forest plots of the lead variant associations with the influenza antibody titers. Blue markers indicate $p < 0.05$. A/H1, influenza A virus subtype H1N1; A/H3, influenza A virus subtype H3N2; B/Vi, influenza virus subtype B/Victoria; B/Ya, influenza virus subtype B/Yamagata.
(G) Manhattan plot for the GWAS of the IgG serostatus in the UKB.
(H) Manhattan plot for the GWAS meta-analysis of the T cell response level against antigen 1.
(I) Manhattan plot for the GWAS meta-analysis of the T cell response level against antigen 2.
The pink horizontal dashed lines in the Manhattan plots denote the genome-wide significance threshold of $p = 5.0 \times 10^{-8}$. Error bars in the forest plots denote 95% confidence intervals. Quantile-quantile plots for the GWASs are shown in Figure S13. vac., vaccination.

the SARS-CoV-2 coronavirus self-test antibody study data of the UKB. This dataset comprises the serostatus (positive or negative) for anti-S-protein IgG antibodies and COVID-19 vaccination status for about 200,000 study participants predominantly of European ancestry. After restricting the participants to those who took the antibody test after the vaccination and excluding those with potential natural infection (STAR Methods), we performed a genome-wide association analysis comparing 59,937 participants with positive IgG vs. 92,969 with negative IgG. The demographic characteristics of the study participants are shown in Table S2. We identified four loci reaching the genome-wide significance (Table S3 and Figure 2G; regional plots shown in Figure S7): the immunoglobulin kappa light chain locus at 2p11 (IGK; $p = 2.3 \times 10^{-8}$ at rs139861548 [2:89246954:C:A]), MHC at 6p21 ($p = 1.9 \times 10^{-33}$ at rs71542430 [6:32630087:T:C]), IGH at 14q32 ($p = 1.5 \times 10^{-9}$ at rs4294714 [14:107160361:A:C]), and 16p11 ($p = 3.2 \times 10^{-8}$ at rs28472312 [16:28826049:C:T]). Of the associated variants, rs4294714 and rs28472312 were found in the imputed genotype data of the Japanese vaccination cohorts. rs4294714 at IGH showed a suggestive association in the Japanese cohorts with a concordant directional effect beyond ancestry ($p = 0.057$; Table S3). The association of rs28472312 at 16p11 was not observed ($p = 0.61$), despite being a common variant in the Japanese cohorts (MAF = 0.30). We note that the A allele of rs139861548 at IGK is rare in East Asian people (e.g., 0.0066 in the East Asian group of gnomAD), which makes genotype imputation challenging in the Japanese dataset. According to the GTEx,[16] rs139861548 is an eQTL variant of *IGKV1-13*, with the A allele upregulating the gene expression in the whole blood ($p = 2.0 \times 10^{-81}$). The serostatus GWAS and expression quantitative trait locus (eQTL) associations showed a high colocalization probability (PP4 = 0.96 by COLOC; Figure S8), suggesting that the causal allele increases gene expression and facilitates antibody production.

Next, to investigate the genetic architecture of the vaccine-induced cellular immunity, we performed an association analysis on the T cell response levels against antigen 1 in the discovery cohort. We identified genome-wide significant associations at MHC ($p = 1.5 \times 10^{-30}$, $\beta = -0.43$, and SE = 0.037 at rs9282156 [6:32633918:C:CAA]), which was replicated in the independent replication cohort ($p = 7.4 \times 10^{-9}$; Table 1; meta-analysis results shown in Figure 2H; a regional plot shown in Figure S9). We also analyzed genetic associations with the T cell response levels against antigen 2 in the discovery and replication cohorts, observing highly concordant association signals ($p = 1.6 \times 10^{-24}$ in the discovery cohort and $p = 3.1 \times 10^{-8}$ in the replication cohort at rs9282156; Figures 2I and S10). The lead variant of the antigen 1 GWAS and antigen 2 GWAS was the same. We note that the SARS-CoV-2-specific IgG antibody titer-associated variant at IGH (rs1043109) did not show association with the T cell response levels against antigen 1 nor 2 ($p = 0.62$ and 0.93).

To investigate longitudinal shifts in the genetic effects on vaccine immunogenicity, we measured the IgG antibody titers in the blood collected at four additional time points (i.e., before the vaccination, 3 and 8 weeks after the vaccination, and 3 months after the vaccination) and the T cell responses at two additional time points (i.e., before the vaccination and 8 weeks after the vaccination) (Figure S11). As expected, none of the vaccine immunogenicity-associated variants showed associations before the vaccination (Figure S12). When we examined the association between each significant locus and the corresponding immunogenicity phenotype, the effect sizes for the MHC-IgG titers, IGH-IgG titers, and MHC-T cell response associations found in the Japanese GWAS were almost consistent across the blood sampling time points after vaccination. In contrast, the effect size of the IGH-IgG serostatus association found in the European GWAS waned over time ($\beta = -0.13$, -0.12, -0.064, and -0.051 at 3 weeks, 8 weeks, 3 months, and 6 months after vaccination, respectively), indicating the importance of considering phenotype measurement timing when designing genetic association studies involving immunogen exposures.

MHC fine-mapping of vaccine immunogenicity-associated HLA variants

Our GWAS revealed the significant contribution of MHC to vaccine-induced immunogenicity, both humoral and cellular immunity. Among the genes densely residing in MHC, sequence variations in the HLA genes are considered to drive most of the association signals. To fine-map the associations within MHC and prioritize candidate causal HLA variants, we applied HLA imputation to the genotype data using a high-resolution HLA reference panel of Japanese ancestry[17] ($n = 1{,}118$; STAR Methods). After applying the stringent post-imputation QC filter, we retained genotype dosages of 64 two-digit, 98 four-field, and 95 six-field HLA alleles and 616 amino acid polymorphisms of the classical HLA genes.

When evaluating the associations of the imputed HLA variants with the IgG antibody titer in the combined dataset of the discovery and replication cohorts, we observed the strongest association signals at the MHC class II region (Figure 3A), in concordance with earlier work of HLA analysis on COVID-19 vaccine immunogenicity in the European populations (Mentzer et al.[2] and Bian et al.[3]). In the MHC class II region, while Mentzer et al. reported HLA-DQB1*06 and Bien et al. reported HLA-DRβ1 amino acid position 71 as the top-associated HLA variant, the top-associated HLA variant in the Japanese dataset was HLA-DRβ1 amino acid position 96 ($p_{omnibus} = 2.1 \times 10^{-14}$). The most common His allele was in strong LD with the lead GWAS variant rs9270585 ($r^2 = 0.96$). When we conditioned on HLA-DQB1*06 or HLA-DRβ1 amino acid position 71, HLA-DRβ1 amino acid position 96 remained significant ($p_{omnibus} = 3.8 \times 10^{-14}$ and 2.2×10^{-8}), demonstrating that the reported HLA variants in the European populations did not explain the association in the Japanese dataset. This difference may be due to the highly population-specific allele frequency spectrum and LD structure of the HLA alleles. Interestingly, among the four amino acid alleles of Gln/Glu/Tyr/His at HLA-DRβ1 amino acid position 96, the most common His allele increased the antibody titer best, and alleles with lower frequency had stronger decreasing effects (Table S4). When we conditioned on HLA-DRβ1 amino acid position 96, no independent association was observed (Figure 3B). HLA-DRβ1 amino acid position 96 was estimated to explain 3.1% of the IgG antibody titer variance.

Next, we tested the associations of the imputed HLA variants with the T cell response against antigen 1. The most strongly

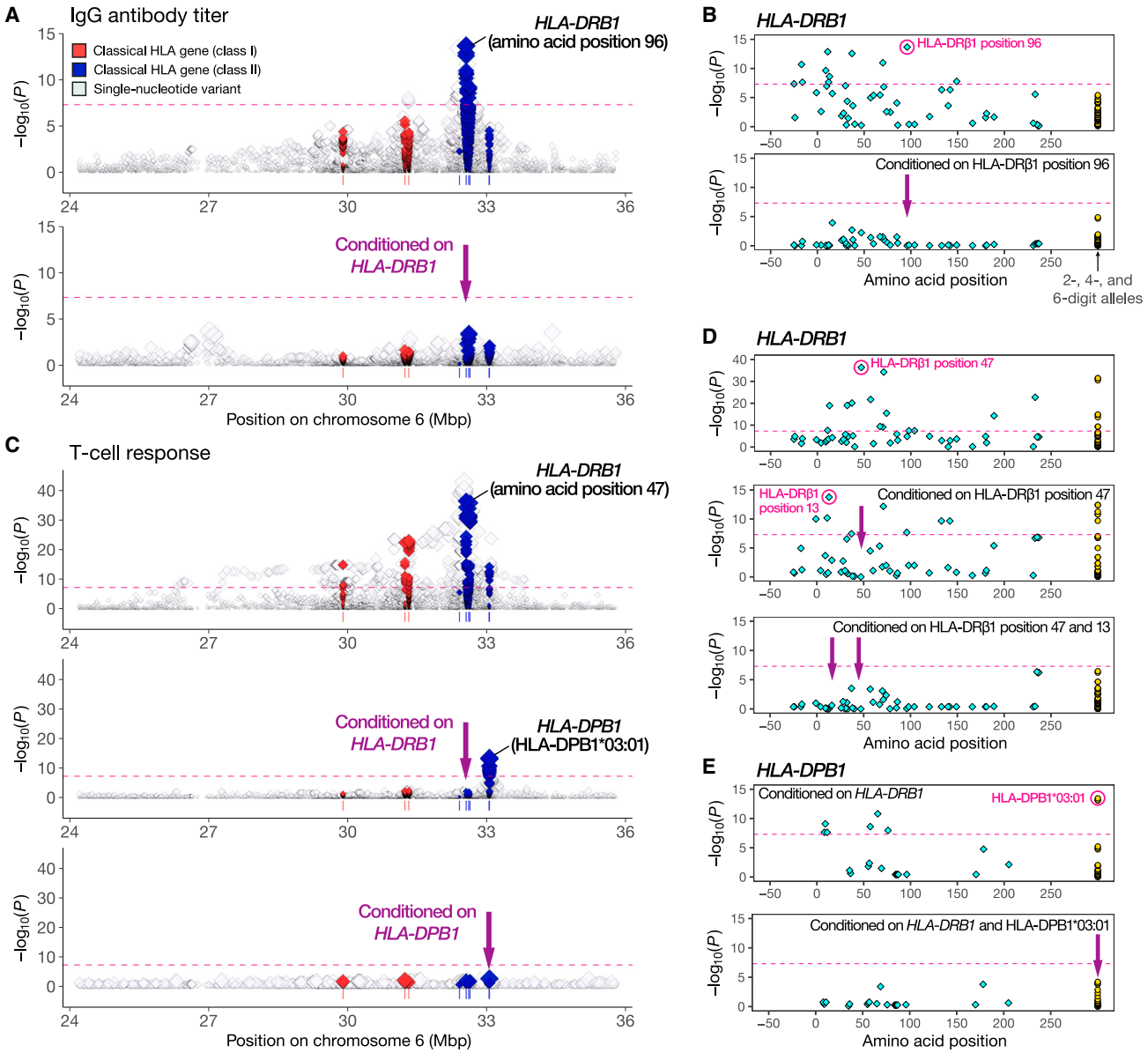


Figure 3. Regional associations of HLA variants with COVID-19 vaccine immunogenicity

(A) Stepwise association analysis of the imputed HLA variants in the MHC region with the IgG antibody titer.

(B) Stepwise association analysis of the imputed amino acid polymorphisms in *HLA-DRB1* with the IgG antibody titer.

(C) Stepwise association analysis of the imputed HLA variants in the MHC region with the T cell response level against antigen 1.

(D and E) Stepwise association analysis of the imputed amino acid polymorphisms in *HLA-DRB1* (D) and *HLA-DPB1* (E) with the T cell response level against antigen 1.

Each diamond represents the $-\log_{10}(p)$ of the variants, including the single-nucleotide variants; two-, four-, and six-digit HLA alleles; and amino acid polymorphisms of the HLA genes in (A) and (C). Each diamond represents the $-\log_{10}(p_{omnibus})$ of the amino acid polymorphisms of the HLA genes in (B), (D), and (E). The pink horizontal dashed lines denote the genome-wide significance threshold of $p = 5.0 \times 10^{-8}$. The most strongly associated amino acid polymorphisms and HLA classical alleles are labeled when their associations showed $p < 5.0 \times 10^{-8}$. Detailed association statistics are available in Table S6.

associated HLA variant was HLA-DRβ1 amino acid position 47, having two alleles of Phe/Tyr ($p_{binary} = 4.3 \times 10^{-37}$; Figure 3C). When conditioning on HLA-DRβ1 amino acid position 47, we observed the top association signal at HLA-DRβ1 amino acid position 13 ($p_{omnibus} = 1.7 \times 10^{-14}$; Figure 3D). After conditioning on HLA-DRβ1 amino acid positions 47 and 13, no other *HLA-DRB1* variants reached the genome-wide significance ($p > 5.0 \times 10^{-8}$). We then sought to identify additional HLA variants modulating the T cell response level independent of *HLA-DRB1*. When we conditioned on all classical *HLA-DRB1* alleles, a significant independent association was found at the classical HLA-DPB1*03:01 allele ($p_{binary} = 3.6 \times 10^{-14}$; Figure 3E). Further conditioning on

HLA-DPB1*03:01 revealed no additional association of *HLA-DPB1* variants. When conditioning on all classical *HLA-DRB1* and *HLA-DPB1* alleles, we observed no other significant associations in the MHC locus (Figure 3E). To explore potential functional mechanisms underlying the T cell response-associated HLA variants, we examined HLA alleles regulating T cell receptor (TCR) chain usage[18] (Figure S14). Among the HLA amino acid positions in high LD with the fine-mapped variants, HLA-DRβ1 position 71, which is in LD with HLA-DRβ1 Arg13 (r^2 = 0.95), showed a high probability of influencing TCR V_α chain usage (probability = 0.99 by Sharon et al.[18]). The bias in TCR usage may lead to the variation in the T cell response.

Finally, we constructed a multivariate full regression model incorporating the identified T cell response-associated HLA variants, which in total explained 14.5% of the T cell response variance (Table S4). HLA-DRβ1 Phe47 demonstrated an increasing effect (β = 0.62, SE = 0.048, $p = 9.1 \times 10^{-36}$) and was the strongest contributor, explaining 8.0% of the total phenotypic variance. For HLA-DRβ1 amino acid position 13, we observed decreasing effects of Arg13 (β = −0.40, SE = 0.065, $p = 1.2 \times 10^{-9}$) and Gly13 (β = −0.23, SE = 0.047, $p = 1.4 \times 10^{-6}$). HLA-DPB1*03:01 showed an increasing effect (β = 0.61, SE = 0.070, $p = 3.4 \times 10^{-18}$). We also performed a stepwise conditional regression analysis using the T cell response level against antigen 2 in the same way, confirming concordant results (Figure S15 and Table S5).

Vaccine immunogenicity-associated alleles modulate the blood proteome

One possible interpretation of our analysis results is that the identified genetic variants modulate immunogenicity against immunogen exposure in general, which is represented by the SARS-CoV-2 vaccines in this study. If this is the case, these variants may affect various molecular phenotypes in the human body, not limited to the specific immunological context of COVID-19 vaccination. We tested this hypothesis by taking the proteomic profile in the circulating blood as a model molecular phenotype. Specifically, we performed protein QTL (pQTL) analysis using proteogenomic datasets from two Japanese cohorts: Biobank Japan[19] (BBJ; *n* = 2,700 individuals) and the Japan COVID-19 Task Force[7,20,21] (JCTF; *n* = 1,536 individuals). Both datasets comprised 2,943 protein levels in the circulating blood measured using the Olink Explore 3072 platform (Figure 4A).

We adopted two distinct analytical approaches separately for the HLA and non-HLA variants. Given the multiple allelic contribution of the HLA variants to the immunogenicity phenotypes, we calculated a genetic score (HLA-GS) for each individual based on the effect sizes estimated by the multivariate regression model, which reflects an overall genetic effect on the phenotype (STAR Methods). When we tested the association of the IgG antibody titer HLA-GS with the protein levels, we identified 15 significant proteins (α = 0.05 by the Bonferroni correction; Figure 4B). We note that 8 of these are encoded outside the MHC locus (i.e., *trans*-association), including immune response-regulating proteins, such as leukocyte immunoglobulin-like receptor subfamily B member 4 (LILRB4) and Fc receptor-like 6 (FCRL6).[22] In contrast, when we examined the association of the T cell response HLA-GS, the strongest associations were primarily observed for the genes encoded within the MHC locus (i.e., *cis*-association), which are potentially driven by the strong and long-range LD within the MHC locus (Figure 4C).

Regarding the non-HLA variants, we performed a standard pQTL analysis assessing the effect of the single variant allele on the protein levels. The IgG antibody titer-associated variant at IGH (14:106208326:G:C) identified in the Japanese GWAS did not show significant associations with the measured protein levels (Figure 4D). The IgG antibody serostatus-associated variant at IGH (14:107160361:A:C) identified in the European GWAS demonstrated three significant *trans*-associations (β = −0.30, SE = 0.024, $p = 6.8 \times 10^{-36}$ for ARHGAP5; β = −0.14, SE = 0.024, $p = 9.4 \times 10^{-9}$ for RGS8; and β = 0.12, SE = 0.024, $p = 6.2 \times 10^{-7}$ for DNAJB8; Figure 4E). The IgG antibody serostatus-associated variant at 16p11 (16:28826049:C:T) identified in the European GWAS had an increasing *cis*-pQTL effect on SULT1A1 (β = 0.15, SE = 0.023, $p = 1.4 \times 10^{-10}$; Figure 4F). These results suggest that the vaccine-induced antibody production-associated variants have widespread immunological effects beyond the context of COVID-19 vaccination, involving changes in the proteomic profile not simply attributable to *cis*-genetic regulation.

Hematopoietic mosaic chromosomal alterations lead to dysregulated immune function

Aging is a well-established factor that reduces vaccine efficacy.[23] Recent studies reported that age-related clonal hematopoiesis, which is commonly observed in apparently healthy elderly individuals, increases the risk of various diseases, including hematological malignancies, cardiovascular diseases, and infectious diseases.[24–28] We hypothesized that clonal hematopoiesis also influences the competency of vaccine-induced antibody production. To test this hypothesis, we analyzed the association of the IgG antibody serostatus data of the UKB with hematopoietic mCAs, which are somatic copy number alterations detected in peripheral leukocytes indicative of clonal hematopoiesis. We specifically focused on expanded mCAs, those present in >10% of peripheral leukocyte DNA, due to their high likelihood of causing phenotypic effects. When we examined expanded autosomal mCA events by adjusting for age, age^2, sex, antibody self-test kit, number of doses, dose intervals, smoking status, Townsend deprivation, body mass index, and the top 10 genetic principal components (PCs), we found the strongest association of any type of mCA event at 14q (Any14q), followed by Any6p (p = 0.0064 and 0.020; Figure 5A).

Of interest, the top-associated chromosome arms of 14q and 6p in the somatic alteration analysis contain the top-associated loci of IGH and MHC in the germline analysis (i.e., GWAS). Motivated by these regionally converging trait associations between germline and somatic mutations, we then restricted the expanded mCA events at 14q and 6p to those affecting IGH (1.1 Mb) and MHC (4.9 Mb) for analysis. For both loci, copy-neutral loss of heterozygosity was the major mCA type compared with loss and gain. IGH-affecting expanded mCA events exhibited consistent decreasing effects on vaccine-induced antibody production (for any type of IGH-affecting expanded mCAs [AnyIGH], p = 0.0048, odds ratio [OR] = 0.34; Figure 5B). We confirmed that the IGH-affecting expanded

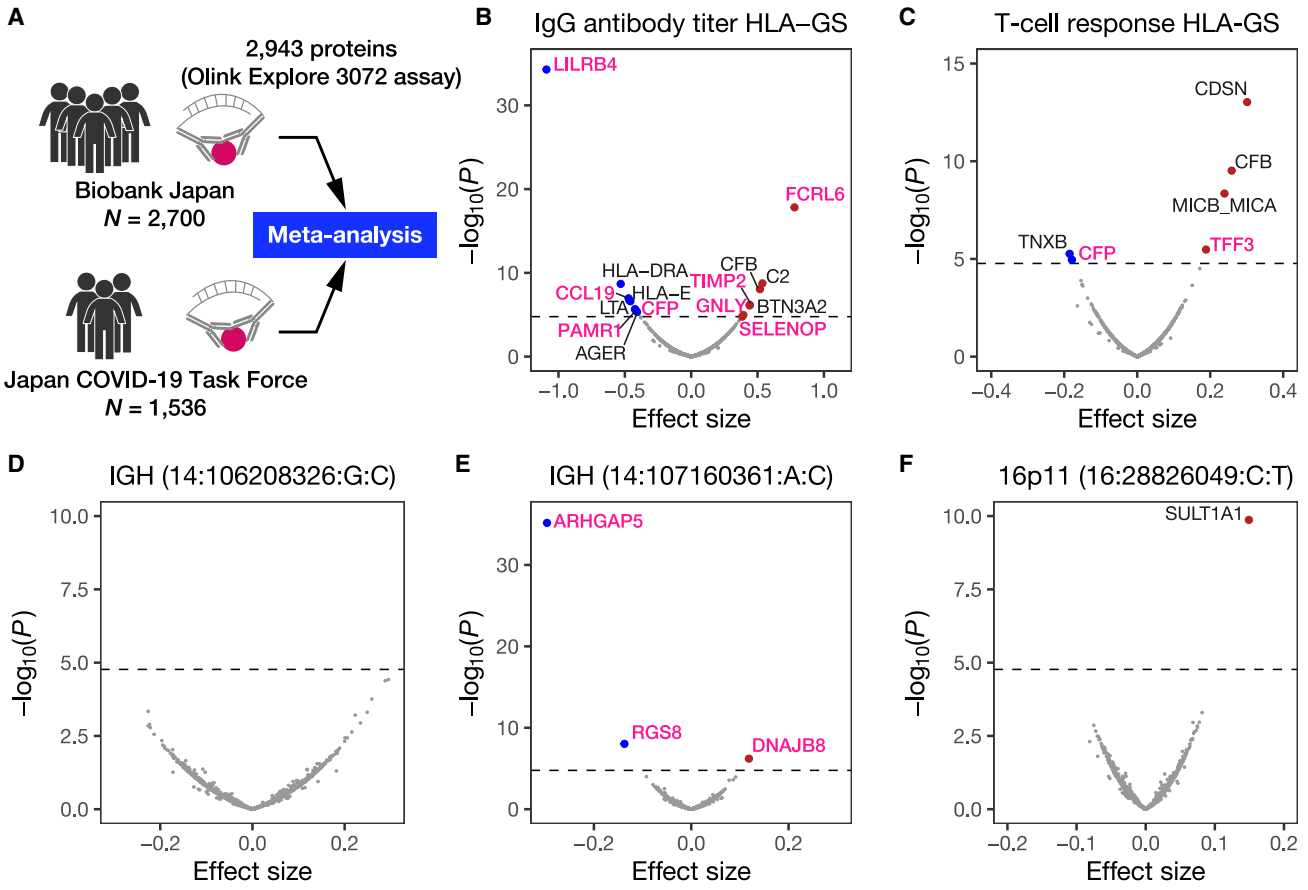


Figure 4. Vaccine immunogenicity-associated alleles modulate blood proteome

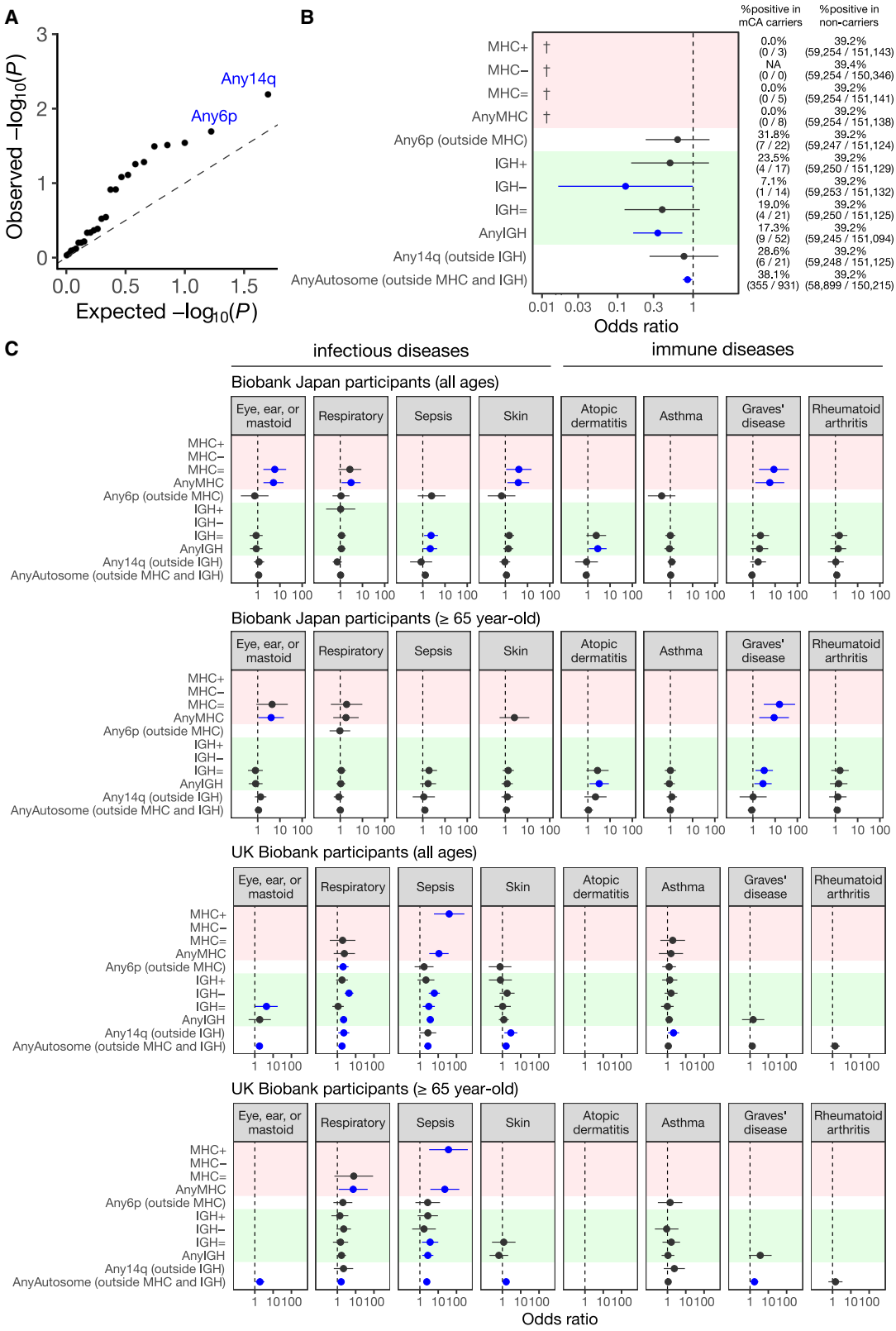
(A) Schematic view of the proteogenomic analysis.

(B–F) Volcano plots representing the associated protein expression in the blood with IgG antibody titer HLA-GRS (B), T cell response HLA-GRS (C), IgG antibody titer-associated variant at IGH found in the Japanese cohorts (D), and IgG serostatus-associated variant at IGH (E) and 16p11 (F) found in the UKB. The x axis and y axis denote the effect sizes and $-\log_{10}(p)$ values in the meta-analysis, respectively. The horizontal dashed lines denote the Bonferroni-corrected significance threshold of $p = 0.05/2{,}943 = 1.7 \times 10^{-5}$. The proteins satisfying the Bonferroni-corrected significance threshold are labeled. *trans*-associated proteins are highlighted in pink, defined as those encoded outside MHC (B) and (C) or outside the 2 Mb window centered around the tested variant (D–F).

mCA association was robustly observed when we excluded individuals having potential immunosuppression-related traits, including cancers, aplastic anemia, neutropenia, chemotherapy, bone marrow or stem cell transplant, and radiotherapy (Figure S16). In contrast, when we tested Any14q mCAs that did not affect IGH (notated as "Any14q (outside IGH)" in Figure 5B), the association was not observed ($p = 0.60$, OR = 0.76). Notably, MHC-affecting expanded mCA carriers were all negative for anti-SARS-CoV-2 IgG at least 1 week after vaccination (0%), whereas 7 of 22 (32%) carriers of Any6p mCAs that did not affect MHC (notated as "Any6p (outside MHC)" in Figure 5B) were IgG positive ($p = 0.34$, OR = 0.62). When we aggregated expanded mCA events throughout the autosomes except for those MHC/IGH affecting, we did not observe such a strong decreasing effect (OR = 0.84), indicating that somatic alterations arising at these germline GWAS-implicated loci specifically contribute to age-related impairment of adaptive immunity.

Given the significant contribution of MHC/IGH-affecting expanded mCAs to dysregulated immune function, we then explored the pleiotropic effect of these classes of mCA events on the infectious and immune diseases in the BBJ (n_{max} = 180,389) and UKB (n_{max} = 482,744), the two large-scale biobanks of Japanese and Europeans, by adjusting for age, age^2, sex, smoking status, and the top 10 genetic PCs (Figure 5C). Given that the pathological backgrounds of infectious and immune diseases are at times different between young and elderly patients, we performed the analysis either using all the individuals or confined to the elderly individuals (≥65 years of age). In the BBJ, 22 MHC-affecting and 198 IGH-affecting expanded mCA carriers were detected; in the UKB, 19 MHC-affecting and 199 IGH-affecting expanded mCA carriers were detected (Figure S17).

MHC-affecting expanded mCA events consistently conferred the risk of infectious diseases of multiple systems in the BBJ (for AnyMHC, $p = 0.0021$, OR = 5.1 with eye, ear, or mastoid; $p = 0.026$, OR = 3.1 with the respiratory system; $p = 0.020$, OR = 3.8 with skin). We observed that MHC-affecting expanded mCA events also conferred the risk of respiratory system



(legend on next page)

infection and sepsis in the elderly (≥65 years of age) in the UKB (for AnyMHC, $p = 0.032$, OR = 7.4 with the respiratory system; $p = 6.8 \times 10^{-4}$, OR = 23.5 with sepsis). Of note, while MHC-affecting expanded mCAs increased the risk of infectious diseases, most Any6p (outside MHC) events, namely the other mCAs at 6p, did not exhibit significant associations, suggesting a locus-specific contribution. IGH-affecting expanded mCA events conferred the risk of sepsis in both the BBJ and the UKB (for AnyIGH, $p = 0.035$, OR = 2.2 in BBJ; $p = 2.3 \times 10^{-8}$, OR = 3.7 in UKB). Strikingly, when we analyzed immune diseases in the elderly, both MHC-affecting and IGH-affecting expanded mCA events increased the risk of Graves' disease in BBJ (for AnyMHC, $p = 0.0051$, OR = 9.0; for AnyIGH, $p = 0.023$, OR = 2.8). On the other hand, the effect of Any14q (outside IGH) was limited (OR = 1.0). While MHC-affecting expanded mCA events were rarely found among the individuals diagnosed with immune diseases in the UKB, IGH-affecting expanded mCAs showed the same directional effect on Graves' disease in the elderly in the UKB (for AnyIGH, $p = 0.079$, OR = 3.6).

In addition to autosomal expanded mCAs, we examined the influence of mLOX in females and mLOY in males, the more common hematopoietic clonal somatic alterations implicated in immune cell function[29] that may explain the sex difference in vaccination responses.[30] In addition to the expanded mLOX/mLOY cutoff of >10%,[27,29] we examined the associations using a cutoff of >5%, another definition of the expanded events.[29] When examining their association with the IgG antibody serostatus, we found that expanded mLOY with cell fraction > 10% showed a negative association ($p = 0.032$; OR = 0.90) and that expanded mLOX did not show a significant association (Figure 6A). We further analyzed the association of expanded mLOX/mLOY with infectious and immune diseases. In general, mLOX showed strong increasing effects, and mLOY showed weak increasing effects on the disease risk. For example, expanded mLOX with cell fraction > 10% conferred Graves' disease ($p = 0.015$, OR = 4.4) and rheumatoid arthritis risk ($p = 0.0014$, OR = 3.8) in the BBJ, particularly in the elderly. Expanded mLOY with cell fraction > 5% conferred asthma risk ($p = 2.0 \times 10^{-10}$, OR = 1.3 in BBJ and $p = 0.044$, OR = 1.1 in the elderly in UKB) and rheumatoid arthritis risk ($p = 0.034$, OR = 1.2 in BBJ and $p = 0.0033$, OR = 1.6 in UKB; Figure 6B).

DISCUSSION

In this study, we established COVID-19 vaccination cohorts of Japanese people and performed GWASs of their vaccine immunogenicity, including SARS-CoV-2-specific IgG antibody titers and T cell responses quantified using multiple assay kits. Our study is one of the first reports on genetic determinants of COVID-19 vaccine immunogenicity in non-European populations. In combination with an analysis of the UKB self-test antibody serostatus dataset, we provided robust evidence of the contribution of genetic variation at the IGH locus to vaccine-induced IgG production beyond ancestry. Of note, we identified a population-specific missense variant in *IGHG1* that strongly decreases SARS-CoV-2-specific IgG antibody titers ($\beta = -0.54$ on the standard normal distribution scale), highlighting the potential importance of taking populations into account in vaccine development and efficacy evaluation.

The IGH locus is among the most structurally diverse and complex loci in the human genome, consisting of an estimated 123–129 variable, 27 diversity, 9 joining, and 5–11 constant gene segments, with extensive copy number variations.[31] Although its pivotal role in the adaptive immune system has made the IGH locus regarded as the promising candidate gene for various infectious and immune diseases, surprisingly, GWAS findings of significant associations of the IGH locus are rather limited to date.[32–34] One possible explanation for this gap is that the structural complexity imposes a technical challenge for the current standard GWAS methodology due to the sparse coverage and poor tagging of this locus by genotyping SNP arrays. Indeed, in our preliminary analysis using only the SNP-array-based genotypes for the genotype imputation, the *IGHG1* missense variant showed moderate imputation performance (imputation *INFO* = 0.36), even though it reached the genome-wide significance threshold (STAR Methods). As we performed in this study, the novel technology of long-read sequencing, as well as traditional Sanger sequencing, should aid in comprehensively and accurately capturing the diversity in this locus.

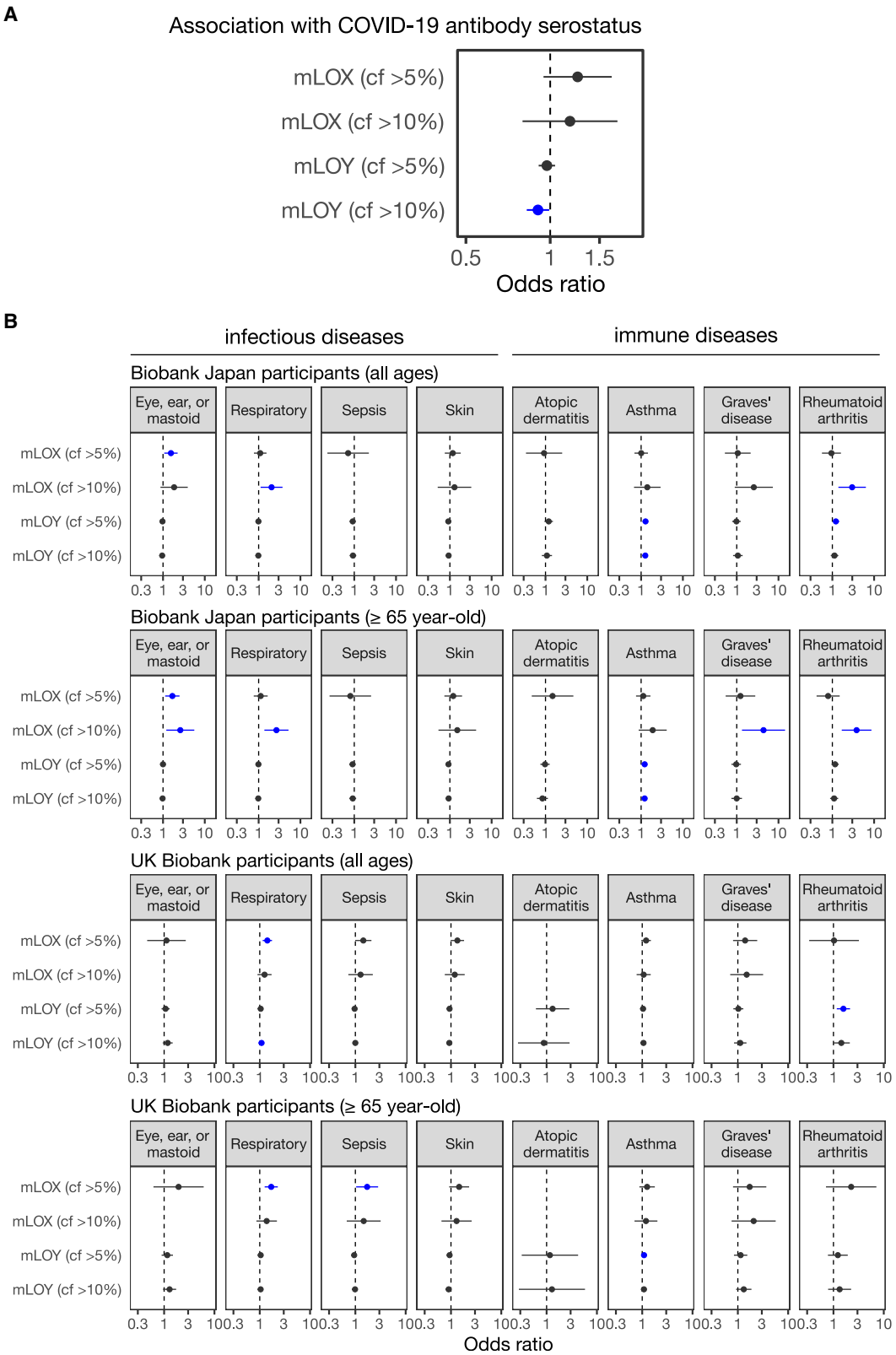
The MHC fine-mapping analysis demonstrated that the MHC class II region drove the association between MHC and antibody production, consistent with previous studies in the European populations.[2,3] The fine-mapped HLA variants were distinct from those identified in the European populations, presumably due to the difference in the allele frequency spectrum and LD structures of the HLA variants shaped by population-specific selective pressure. We further demonstrated that MHC is also a strong genetic determinant of the vaccine-induced T cell immune response. The fine-mapped HLA variants were different between the antibody titers and the T cell responses, which may reflect the difference in the underlying immunological processes in which the MHC molecule plays the antigen-presentation role between humoral and cellular immunity. The T cell response-associated amino acid residues, HLA-DRβ1 positions 13 and 47, were located in the functional pocket of the MHC molecule, suggesting their functional influence on antigen-presentation ability (Figure S18). We note that some of the fine-mapped HLA variants in our study were previously associated with other pathogen titers. HLA-DRβ1 position

Figure 5. Associations of hematopoietic expanded mCAs at MHC/IGH with immune dysregulation

(A) Quantile-quantile plot of the p values of the autosomal mCA class associations with the COVID-19 IgG antibody serostatus in the UKB. The x axis denotes $-\log_{10}(p)$ expected from a uniform distribution. The y axis denotes the observed $-\log_{10}(p)$. The top two strongly associated mCA classes are labeled.

(B) Forest plot of the MHC/IGH-affecting mCA associations with the COVID-19 IgG antibody serostatus in the UKB. Blue markers indicate $p < 0.05$. Daggers (†) indicate that there were no seropositive mCA carriers.

(C) Forest plots of the MHC/IGH-affecting mCA associations with the infectious or immune diseases in the BBJ and UKB in all age groups or the elderly group (≥65 years of age). Association test results are shown where at least two cases carrying the mCA class were found. Error bars in the forest plots denote 95% confidence intervals. Blue markers indicate $p < 0.05$. +, gain; −, loss; =, copy-neutral loss of heterozygosity.



(legend on next page)

96 was reported as associated with influenza A antibody serostatus.[35] One of the HLA variants associated with T cell response in our study, HLA-DRβ1 position 13, was reported as associated with Merkel cell polyomavirus antibody serostatus.[35] Moreover, another HLA variant associated with T cell response in our study, HLA-DPB1*03:01, was reported as associated with rubella IgG levels.[36] These previous reports corroborate the functional significance of the fine-mapped HLA amino acid residues and alleles.

Hematopoietic expanded mCA events affecting the vaccine immunogenicity-associated loci implicated by the GWAS, namely MHC and IGH, also impaired vaccine-induced IgG production. While mCAs affecting specific chromosome arms were implicated in subsequent hematological malignancy development,[24,25,28] to our knowledge, this study is the first to reveal the region-specific contribution of mCAs to non-malignant traits rather than genome-wide mCA aggregation. Furthermore, we observed that IGH-affecting mCAs conferred risk of Graves' disease, an autoimmune disease characterized by the presence of thyroid-specific autoantibodies causing hyperthyroidism. We also demonstrated the impact of the more common and sex-specific hematopoietic clonal somatic alterations, mLOX and mLOY, on the immune-related phenotypes. The role of clonal hematopoiesis in the development of autoimmune diseases has been a controversial topic.[37–39] Our results provide a perspective that explains this controversy by focusing on the somatic alterations rising at the immunologically important loci.

Limitations of the study

We note that mCA events restricted to those affecting specific genomic loci tend to be rare and limit the statistical power. For this reason, a limitation of this study is that some of the mCA associations with infectious and immune diseases might not reach statistical significance if we consider the potential multiple testing. Due to data availability, mCA associations with vaccine-induced antibody production were analyzed only in the single population of the UKB. Further investigations in other populations would be needed to evaluate the generalizability of our findings. As with other observational studies of mCA, the examination of molecular mechanisms mediating the mCA effects on immune function using experimental models would be warranted as future work.

RESOURCE AVAILABILITY

Lead contact

Requests for further information should be directed to the lead contact, Yukinori Okada (yuki-okada@m.u-tokyo.ac.jp).

Materials availability

The materials that support the findings of this study are available from the corresponding authors upon reasonable request. Please contact the lead contact, Yukinori Okada (yuki-okada@m.u-tokyo.ac.jp) for additional information.

Data and code availability

GWAS summary statistics are available upon request to the lead contact. HLA fine-mapping summary statistics are provided as Table S6. The SNP array data and GWAS genotype data of the BBJ are available at the NBDC Human Database under accession no. hum0014 (https://humandbs.dbcls.jp/en/hum0014-latest). The SNP array data of the JCTF are available at the European Genome-Phenome Archive (EGA) under accession no. EGAS00001006284 (https://ega-archive.org/studies/EGAS00001006284). The UKB analysis was conducted via application no. 47821 (https://www.ukbiobank.ac.uk). This paper does not report original code. Any additional information required to reanalyze the data reported in this work is available from the lead contact upon reasonable request.

ACKNOWLEDGMENTS

We would like to thank all the participants involved in this study. We sincerely thank Dr. Michiaki Kubo and the members of the RIKEN Center for Integrative Medical Sciences for their support of this study. We thank all the participants involved in this study and all the members of the JCTF for their support. We thank J. Kitano, the e-Parcel Corporation, and the Ascend Corporation for supporting the JCTF. This research was supported by the KAKENHI Grants-in-Aid from the Japanese Society for the Promotion of Science (JSPS) (grant nos. 23K14451 [to K.S.] and 22H00476 [to Y. Okada]); the Japan Agency for Medical Research and Development (AMED) (grant nos. JP23fk0108686 [to Y.U.]; JP21fk0108469 and JP24tm0524008 [to H.N.]; and JP24km0405217, JP24ek0109594, JP24ek0410113, JP24kk0305022, JP243fa627002, JP243fa627010, JP243fa627011, JP24zf0127008, JP24tm0524002, JP24wm0625504, and JP24gm1810011 [to Y. Okada]); JST Moonshot R&D (grant nos. JPMJMS2021 and JPMJMS2024 [to Y. Okada]); JST PRESTO (grant nos. JPMJPR21R7 [to H.N.], JPMJPR22R2 [to Y.U.], Keio University Academic Development Fund [Joint Research] [to H.N.], the Japan Prize Foundation [to H.N.], and Public Foundation of Vaccination Research Center [to Y.U.]; the Takeda Science Foundation; the Ono Pharmaceutical Foundation for Oncology, Immunology, and Neurology; the Bioinformatics Initiative of Osaka University Graduate School of Medicine; the Institute for Open and Transdisciplinary Research Initiatives; the Center for Infectious Disease Education and Research (CiDER); the Center for Advanced Modality and DDS (CAMaD), Osaka University; and RIKEN Advanced General Intelligence for Science Program (TRIP-AGIS).

AUTHOR CONTRIBUTIONS

K.S., Y.U., Y. Okada, and H.N. designed the study. K.S., Y.U., R.S., S.N., T. Nakanishi, T. Naito, G.S., M. Kanai, A.L., Y. Okada, and H.N. analyzed the data. K.S. and Y.U. wrote the manuscript with input from all authors. Y. Okada and H.N. reviewed and edited the manuscript. N.H., M.M., H.M., S.O., Y. Okada, and H.N. supervised the study. All authors contributed to the generation of the primary data incorporated in the study.

DECLARATION OF INTERESTS

Q.S.W. is an employee of Calico Life Sciences LLC.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

Figure 6. Associations of hematopoietic expanded mLOX/mLOY with immune dysregulation

(A) Forest plot of the expanded mLOX/mLOY associations with the COVID-19 IgG antibody serostatus in the UKB. Blue markers indicate $p < 0.05$.

(B) Forest plots of the expanded mLOX/mLOY associations with the infectious or immune diseases in the BBJ and UKB in all age groups or the elderly group (≥65 years of age).

Association test results are shown where at least two cases carrying the mCA class were found. Error bars in the forest plots denote 95% confidence intervals. Blue markers indicate $p < 0.05$. cf, cell fraction.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2025.100783.

Received: July 2, 2024
Revised: November 27, 2024
Accepted: February 6, 2025
Published: March 4, 2025

REFERENCES

1. Zimmermann, P., and Curtis, N. (2019). Factors That Influence the Immune Response to Vaccination. Clin. Microbiol. Rev. *32*, e00084-18. https://doi.org/10.1128/CMR.00084-18.
2. Mentzer, A.J., O'Connor, D., Bibi, S., Chelysheva, I., Clutterbuck, E.A., Demissie, T., Dinesh, T., Edwards, N.J., Felle, S., Feng, S., et al. (2023). Human leukocyte antigen alleles associate with COVID-19 vaccine immunogenicity and risk of breakthrough infection. Nat. Med. *29*, 147–157. https://doi.org/10.1038/s41591-022-02078-6.
3. Bian, S., Guo, X., Yang, X., Wei, Y., Yang, Z., Cheng, S., Yan, J., Chen, Y., Chen, G.-B., Du, X., et al. (2024). Genetic determinants of IgG antibody response to COVID-19 vaccination. Am. J. Hum. Genet. *111*, 181–199. https://doi.org/10.1016/j.ajhg.2023.12.005.
4. Karlsson, E.K., Kwiatkowski, D.P., and Sabeti, P.C. (2014). Natural selection and infectious disease in human populations. Nat. Rev. Genet. *15*, 379–393. https://doi.org/10.1038/nrg3734.
5. Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.-C., Freiman, A., Sams, A.J., Hebert, S., et al. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell *167*, 657–669.e21. https://doi.org/10.1016/j.cell.2016.09.025.
6. Kanai, M., Andrews, S.J., Cordioli, M., Stevens, C., Neale, B.M., Daly, M., Ganna, A., Pathak, G.A., Iwasaki, A., Karjalainen, J., et al. (2023). A second update on mapping the human genetic architecture of COVID-19. Nature *621*, E7–E26. https://doi.org/10.1038/s41586-023-06355-3.
7. Namkoong, H., Edahiro, R., Takano, T., Nishihara, H., Shirai, Y., Sonehara, K., Tanaka, H., Azekawa, S., Mikami, Y., Lee, H., et al. (2022). DOCK2 is involved in the host genetics and biology of severe COVID-19. Nature *609*, 754–760. https://doi.org/10.1038/s41586-022-05163-5.
8. Kakiuchi, N., and Ogawa, S. (2021). Clonal expansion in non-cancer tissues. Nat. Rev. Cancer *21*, 239–256. https://doi.org/10.1038/s41568-021-00335-3.
9. Bolton, K.L., Koh, Y., Foote, M.B., Im, H., Jee, J., Sun, C.H., Safonov, A., Ptashkin, R., Moon, J.H., Lee, J.Y., et al. (2021). Clonal hematopoiesis is associated with risk of severe Covid-19. Nat. Commun. *12*, 5975. https://doi.org/10.1038/s41467-021-26138-6.
10. Sano, S., Horitani, K., Ogawa, H., Halvardson, J., Chavkin, N.W., Wang, Y., Sano, M., Mattisson, J., Hata, A., Danielsson, M., et al. (2022). Hematopoietic loss of Y chromosome leads to cardiac fibrosis and heart failure mortality. Science *377*, 292–297. https://doi.org/10.1126/science.abn3100.
11. Namba, S., Saito, Y., Kogure, Y., Masuda, T., Bondy, M.L., Gharahkhani, P., Gockel, I., Heider, D., Hillmer, A., Jankowski, J., et al. (2023). Common Germline Risk Variants Impact Somatic Alterations and Clinical Features across Cancers. Cancer Res. *83*, 20–27. https://doi.org/10.1158/0008-5472.CAN-22-1492.
12. Perkmann, T., Mucher, P., Ösze, D., Müller, A., Perkmann-Nagele, N., Koller, T., Radakovics, A., Flieder, I., Repl, M., Marculescu, R., et al. (2023). Comparison of five Anti-SARS-CoV-2 antibody assays across three doses of BNT162b2 reveals insufficient standardization of SARS-CoV-2 serology. J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol. *158*, 105345. https://doi.org/10.1016/j.jcv.2022.105345.
13. Vogrig, M., Berger, A.-E., Bourlet, T., Waeckel, L., Haccourt, A., Chanavat, A., Hupin, D., Roche, F., Botelho-Nevers, E., Pozzetto, B., and Paul, S. (2023). Monitoring of Both Humoral and Cellular Immunities Could Early Predict COVID-19 Vaccine Efficacy Against the Different SARS-CoV2 Variants. J. Clin. Immunol. *43*, 31–45. https://doi.org/10.1007/s10875-022-01354-x.
14. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.
15. Chen, S., Francioli, L.C., Goodrich, J.K., Collins, R.L., Kanai, M., Wang, Q., Alföldi, J., Watts, N.A., Vittal, C., Gauthier, L.D., et al. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. Nature *625*, 92–100. https://doi.org/10.1038/s41586-023-06045-0.
16. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
17. Hirata, J., Hosomichi, K., Sakaue, S., Kanai, M., Nakaoka, H., Ishigaki, K., Suzuki, K., Akiyama, M., Kishikawa, T., Ogawa, K., et al. (2019). Genetic and phenotypic landscape of the major histocompatibilty complex region in the Japanese population. Nat. Genet. *51*, 470–480. https://doi.org/10.1038/s41588-018-0336-0.
18. Sharon, E., Sibener, L.V., Battle, A., Fraser, H.B., Garcia, K.C., and Pritchard, J.K. (2016). Genetic variation in MHC proteins is associated with T cell receptor expression biases. Nat. Genet. *48*, 995–1002. https://doi.org/10.1038/ng.3625.
19. Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., et al. (2017). Overview of the BioBank Japan Project: Study design and profile. J. Epidemiol. *27*, S2–S8. https://doi.org/10.1016/j.je.2016.12.005.
20. Wang, Q.S., Edahiro, R., Namkoong, H., Hasegawa, T., Shirai, Y., Sonehara, K., Tanaka, H., Lee, H., Saiki, R., Hyugaji, T., et al. (2022). The whole blood transcriptional regulation landscape in 465 COVID-19 infected samples from Japan COVID-19 Task Force. Nat. Commun. *13*, 4830. https://doi.org/10.1038/s41467-022-32276-2.
21. Wang, Q.S., Hasegawa, T., Namkoong, H., Saiki, R., Edahiro, R., Sonehara, K., Tanaka, H., Azekawa, S., Chubachi, S., Takahashi, Y., et al. (2024). Statistically and functionally fine-mapped blood eQTLs and pQTLs

from 1,405 humans reveal distinct regulation patterns and disease relevance. Nat. Genet. *56*, 2054–2067. https://doi.org/10.1038/s41588-024-01896-3.

22. Krishna, C., Chiou, J., Sakaue, S., Kang, J.B., Christensen, S.M., Lee, I., Aksit, M.A., Kim, H.I., von Schack, D., Raychaudhuri, S., et al. (2024). The influence of HLA genetic variation on plasma protein expression. Nat. Commun. *15*, 6469. https://doi.org/10.1038/s41467-024-50583-8.
23. Collier, D.A., Ferreira, I.A.T.M., Kotagiri, P., Datir, R.P., Lim, E.Y., Touizer, E., Meng, B., Abdullahi, A., , et al.; CITIID-NIHR BioResource COVID-19 Collaboration; Elmer, A., Kingston, N. (2021). Age-related immune response heterogeneity to SARS-CoV-2 vaccine BNT162b2. Nature *596*, 417–422. https://doi.org/10.1038/s41586-021-03739-1.
24. Loh, P.-R., Genovese, G., Handsaker, R.E., Finucane, H.K., Reshef, Y.A., Palamara, P.F., Birmann, B.M., Talkowski, M.E., Bakhoum, S.F., McCarroll, S.A., and Price, A.L. (2018). Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. Nature *559*, 350–355. https://doi.org/10.1038/s41586-018-0321-x.
25. Loh, P.-R., Genovese, G., and McCarroll, S.A. (2020). Monogenic and polygenic inheritance become instruments for clonal selection. Nature *584*, 136–141. https://doi.org/10.1038/s41586-020-2430-6.
26. Terao, C., Suzuki, A., Momozawa, Y., Akiyama, M., Ishigaki, K., Yamamoto, K., Matsuda, K., Murakami, Y., McCarroll, S.A., Kubo, M., et al. (2020). Chromosomal alterations among age-related haematopoietic clones in Japan. Nature *584*, 130–135. https://doi.org/10.1038/s41586-020-2426-2.
27. Zekavat, S.M., Lin, S.-H., Bick, A.G., Liu, A., Paruchuri, K., Wang, C., Uddin, M.M., Ye, Y., Yu, Z., Liu, X., et al. (2021). Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of infection. Nat. Med. *27*, 1012–1024. https://doi.org/10.1038/s41591-021-01371-0.
28. Saiki, R., Momozawa, Y., Nannya, Y., Nakagawa, M.M., Ochi, Y., Yoshizato, T., Terao, C., Kuroda, Y., Shiraishi, Y., Chiba, K., et al. (2021). Combined landscape of single-nucleotide variants and copy number alterations in clonal hematopoiesis. Nat. Med. *27*, 1239–1249. https://doi.org/10.1038/s41591-021-01411-9.
29. Liu, A., Genovese, G., Zhao, Y., Pirinen, M., Zekavat, S.M., Kentistou, K.A., Yang, Z., Yu, K., Vlasschaert, C., Liu, X., et al. (2024). Genetic drivers and cellular selection of female mosaic X chromosome loss. Nature *631*, 134–141. https://doi.org/10.1038/s41586-024-07533-7.
30. Flanagan, K.L., Fink, A.L., Plebanski, M., and Klein, S.L. (2017). Sex and Gender Differences in the Outcomes of Vaccination over the Life Course. Annu. Rev. Cell Dev. Biol. *33*, 577–599. https://doi.org/10.1146/annurev-cellbio-100616-060718.
31. Watson, C.T., Steinberg, K.M., Huddleston, J., Warren, R.L., Malig, M., Schein, J., Willsey, A.J., Joy, J.B., Scott, J.K., Graves, T.A., et al. (2013). Complete Haplotype Sequence of the Human Immunoglobulin Heavy-Chain Variable, Diversity, and Joining Genes and Characterization of Allelic and Copy-Number Variation. Am. J. Hum. Genet. *92*, 530–546. https://doi.org/10.1016/j.ajhg.2013.03.004.
32. Parks, T., Mirabel, M.M., Kado, J., Auckland, K., Nowak, J., Rautanen, A., Mentzer, A.J., Marijon, E., Jouven, X., Perman, M.L., et al. (2017). Association between a common immunoglobulin heavy chain allele and rheumatic heart disease risk in Oceania. Nat. Commun. *8*, 14946. https://doi.org/10.1038/ncomms14946.
33. 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. https://doi.org/10.1038/ng.3897.
34. Morii, W., Kasai, K., Nakamura, T., Hayashi, D., Hara, M., Naito, T., Sonehara, K., Fukuie, T., Saito-Abe, M., Yang, L., et al. (2023). A genome-wide association study for allergen component sensitizations identifies allergen component–specific and allergen protein group–specific associations. J. Allergy Clin. Immunol. Glob. *2*, 100086. https://doi.org/10.1016/j.jacig.2023.100086.
35. 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. https://doi.org/10.1016/j.ajhg.2015.09.008.
36. 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. https://doi.org/10.1186/s13073-018-0568-8.
37. Savola, P., Lundgren, S., Keränen, M.A.I., Almusa, H., Ellonen, P., Leirisalo-Repo, M., Kelkka, T., and Mustjoki, S. (2018). Clonal hematopoiesis in patients with rheumatoid arthritis. Blood Cancer J. *8*, 1–5. https://doi.org/10.1038/s41408-018-0107-2.
38. Hecker, J.S., Hartmann, L., Rivière, J., Buck, M.C., van der Garde, M., Rothenberg-Thurley, M., Fischer, L., Winter, S., Ksienzyk, B., Ziemann, F., et al. (2021). CHIP and hips: clonal hematopoiesis is common in patients undergoing hip arthroplasty and is associated with autoimmune disease. Blood *138*, 1727–1732. https://doi.org/10.1182/blood.2020010163.
39. Hiitola, E., Korhonen, J., Koskela, J., Kankainen, M., Alakuijala, M., Kokkonen, H., Liu, A., Häppölä, P., Savola, P., Kelkka, T., et al. (2023). Clonal Hematopoiesis Associated with Rheumatoid Arthritis. Blood *142*, 2698. https://doi.org/10.1182/blood-2023-173664.
40. Altshuler, D., and Donnelly, P.; The International HapMap Consortium (2005). A haplotype map of the human genome. Nature *437*, 1299–1320. https://doi.org/10.1038/nature04226.
41. Tadaka, S., Hishinuma, E., Komaki, S., Motoike, I.N., Kawashima, J., Saigusa, D., Inoue, J., Takayama, J., Okamura, Y., Aoki, Y., et al. (2021). jMorp updates in 2020: large enhancement of multi-omics data resources on the general Japanese population. Nucleic Acids Res. *49*, D536–D544. https://doi.org/10.1093/nar/gkaa1034.
42. 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. https://doi.org/10.1186/s13742-015-0047-8.
43. Delaneau, O., Zagury, J.-F., Robinson, M.R., Marchini, J.L., and Dermitzakis, E.T. (2019). Accurate, scalable and integrative haplotype estimation. Nat. Commun. *10*, 5436. https://doi.org/10.1038/s41467-019-13225-y.
44. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287. https://doi.org/10.1038/ng.3656.
45. Naito, T., Suzuki, K., Hirata, J., Kamatani, Y., Matsuda, K., Toda, T., and Okada, Y. (2021). A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. Nat. Commun. *12*, 1639. https://doi.org/10.1038/s41467-021-21975-x.
46. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. Nat. Genet. *53*, 1097–1103. https://doi.org/10.1038/s41588-021-00870-7.
47. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191. https://doi.org/10.1093/bioinformatics/btq340.
48. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. *10*, e1004383. https://doi.org/10.1371/journal.pgen.1004383.
49. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

50. Uwamino, Y., Yokoyama, T., Sato, Y., Shibata, A., Kurafuji, T., Tanabe, A., Noguchi, M., Arai, T., Ohno, A., Yokota, H., et al. (2023). Humoral and cellular immune response dynamics in Japanese healthcare workers up to six months after receiving a third dose of BNT162b2 monovalent vaccine. Vaccine *41*, 1545–1549. https://doi.org/10.1016/j.vaccine.2023.01.049.

51. Sonehara, K., Kimura, Y., Nakano, Y., Ozawa, T., Takahashi, M., Suzuki, K., Fujii, T., Matsushita, Y., Tomiyama, A., Kishikawa, T., et al. (2022). A common deletion at BAK1 reduces enhancer activity and confers risk of intracranial germ cell tumors. Nat. Commun. *13*, 4478. https://doi.org/10.1038/s41467-022-32005-9.

52. Okada, Y., Momozawa, Y., Sakaue, S., Kanai, M., Ishigaki, K., Akiyama, M., Kishikawa, T., Arai, Y., Sasaki, T., Kosaki, K., et al. (2018). Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. Nat. Commun. *9*, 1631. https://doi.org/10.1038/s41467-018-03274-0.

53. Sonehara, K., Sakaue, S., Maeda, Y., Hirata, J., Kishikawa, T., Yamamoto, K., Matsuoka, H., Yoshimura, M., Nii, T., Ohshima, S., et al. (2022). Genetic architecture of microRNA expression and its link to complex diseases in the Japanese population. Hum. Mol. Genet. *31*, 1806–1820. https://doi.org/10.1093/hmg/ddab361.

54. 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. https://doi.org/10.1038/s41586-018-0579-z.

55. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873. https://doi.org/10.1093/bioinformatics/btq559.

56. Okada, Y., Momozawa, Y., Ashikawa, K., Kanai, M., Matsuda, K., Kamatani, Y., Takahashi, A., and Kubo, M. (2015). Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. Nat. Genet. *47*, 798–802. https://doi.org/10.1038/ng.3310.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human whole blood	This study	N/A
Human DNA extracted from blood	This study	N/A
Critical commercial assays		
SARS-CoV-2 IgG II Quant assay	Abbott	6S60
Ortho VITROS SARS-CoV-2 spike-specific quantitative IgG assay	Ortho	N/A
QuantiFERON SARS-CoV-2 assay	QIAGEN	626725
Deposited data		
HapMap individuals genotype data	The International HapMap Consortium[40]	https://ftp.ncbi.nlm.nih.gov/hapmap/
Allele frequency reference panel of Tohoku Medical Megabank Project	Tadaka et al.[41]	https://jmorp.megabank.tohoku.ac.jp/downloads
UK Biobank Genotype and phenotype data	UK Biobank	https://www.ukbiobank.ac.uk/
HLA imputation reference panel	Hirata et al.[17]	National Bioscience Database Center (NBDC) Human Database (https://humandbs.biosciencedbc.jp/en/) with the accession ID hum0114, which is available through application at https://humandbs.biosciencedbc.jp/en/hum0114-latest
BioBank Japan	Nagai et al.[19]	Japanese Genotype-phenotype Archive (JGA) with the accession ID JGAD000836, which is available through application at https://humandbs.biosciencedbc.jp/en/hum0014-latest
Japan COVID-19 Task Force	Wang et al.[21]	European Genome-Phenome Archive (EGA) with the accession ID EGAS00001006284
UK biobank hematopoietic mosaic chromosomal alteration data	Loh et al.[25]	https://www.ukbiobank.ac.uk/ (Return 3094)
Software and algorithms		
PLINK2	Chang et al.[42]	https://www.cog-genomics.org/plink/2.0/
SHAPEIT4	Delaneau et al.[43]	https://github.com/odelaneau/shapeit4
Minimac4	Das et al.[44]	https://github.com/statgen/Minimac4
DEEP*HLA	Naito et al.[45]	https://github.com/tatsuhikonaito/DEEP-HLA
REGENIE	Mbatchou et al.[46]	https://rgcgithub.github.io/regenie/
METAL	Willer et al.[47]	https://github.com/statgen/METAL
pbmm2	Pacific Biosciences	https://github.com/PacificBiosciences/pbmm2
COLOC	Giambartolomei et al.[48]	https://chr1swallace.github.io/coloc/
R	The R Foundation for Statistical Computing	https://www.r-project.org
SMRT Link	Pacific Biosciences	https://www.pacb.com
OlinkAnalyze	Olink Proteomics	https://github.com/Olink-Proteomics/OlinkRPackage
MoChA	Loh et al.[24]	https://github.com/freeseek/mocha
Integrative Genomics Viewer	Robinson et al.[49]	https://igv.org/

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study populations

This study was conducted with the approval of the ethics committee of Keio University and Osaka University. All participants provided written informed consent after being given a full explanation of the purpose of the study and the methods to be employed.

Discovery cohort

The discovery cohort (trial registration number: UMIN000046185) consisted of university staff and students of Keio University Hiyoshi Campus (Yokohama, Japan) and Mita Campus (Tokyo, Japan). The staff and students received two doses of the mRNA-1273 bivalent vaccine (Spikevax Intramuscular Injection, Moderna, Cambridge, USA) by mass vaccination performed at Keio University Mita Campus between June 21th and September 15th, 2021. Almost six months after the second dose (January 11th to February 4th, 2022), they were recruited to the study. Blood samples were collected from all participants just after inclusion. Information about age, sex, and date of vaccination was collected using a questionnaire and validated using the mass vaccination record. We examined the participants' past medical histories and excluded 34 participants from the analysis who had a history of COVID-19.

Replication cohort

The replication cohort (trial registration number: UMIN000043340) consisted of healthcare workers and university staff of Keio University Shinanomachi Campus (Tokyo, Japan) recruited between February 26th and March 9th, 2021[50]. All participants were recruited before vaccination, and after inclusion, they received two doses of the BNT162b2 monovalent vaccine (COMIRNATY intramuscular injection, Pfizer, New York, USA) by mass vaccination performed between March 1st and April 6th. Blood sample collections of all participants were performed at the following time points: before the vaccination, three and eight weeks after the second dose, and three and six months after the second dose. Information about age and sex was collected before the vaccination using a questionnaire. Information about the date of vaccination was obtained using a web-form three weeks after the second dose. The age, sex, and date of vaccination were validated using the mass vaccination record. We examined the participants' past medical histories and excluded 7 participants from the analysis who had a history of COVID-19.

METHOD DETAILS

COVID-19 vaccine immunogenicity measurement

IgG antibody titer

We quantitatively measured IgG antibody titers against the SARS-CoV-2 receptor-binding domain spike protein in all serum samples from the discovery and replication cohorts, using SARS-CoV-2 IgG II Quant reagents and an Alinity Analyzer (Abbott Laboratories, Abbott Park, IL, USA), according to the manufacturer's instructions (assay kit 1). Furthermore, as an independent assay method for technical validation, we quantitatively measured IgG antibody titers against the S1 subunit of the spike protein of SARS-CoV-2 in the serum samples from the discovery cohort, using Ortho VITROS SARS-CoV-2 spike-specific quantitative IgG assay (Ortho Clinical Diagnostics, New Jersey, USA), according to the manufacturer's instructions (assay kit 2).

T cell response

To evaluate the antigen-specific T-cell responses, we performed QuantiFERON SARS-CoV-2 tests (QFN) on the discovery and replication cohorts. Whole blood samples were collected into lithium heparin tubes before vaccination, eight weeks and six months after the second dose. The collected samples were transferred to four QuantiFERON SARS-CoV-2 tubes (Qiagen, Hilden, Germany) coated with antigen 1, antigen 2, phytohemagglutinin (positive control), and no peptide (negative control). Antigen 1 is an epitope of $CD4^+$ T-cells derived from the S1 subunit and antigen 2 is an epitope of $CD4^+$ and $CD8^+$T-cells derived from the S1 and S2 subunits. The tubes were incubated at 37 °C for 16–24 h, and then the samples were tested using the QuantiFERON SARS-CoV-2 ELISA kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, and the DS2 fully automated microplate enzyme immunoassay reader (Dynex Technologies, Chantilly, VA, USA). Interferon-gamma levels for antigen 1 (IFN-γ for Ag1) and antigen 2 (IFN-γ for Ag2) were calculated correcting for the negative control value.

Influenza virus antibody titer measurement

All the participants received influenza vaccination by workplace vaccination performed between October 6th and November 22nd, 2021. Antibody titers against influenza virus were measured in the serum samples from the replication cohort by the hemagglutination inhibition (HAI) assay. HAI assay was performed using Influenza Virus HI Reagents SEIKEN (Denka Co Ltd, Tokyo, Japan). The HAI assay is the most utilized canonical method to quantify influenza-specific antibodies for influenza vaccination clinical studies. HAI assays were conducted using hemagglutinin antigens and antisera for influenza vaccine strains of the 2022/2023 season: A/Victoria/1/2020 (H1N1), A/Darwin/9/2021 (H3N2), B/Phuket/3073/2013 (B/Yamagata), and B/Austria/1359417/2021 (B/Victoria). Serum samples were diluted in a range from 10- to 1280-fold, then mixed with chicken red blood cells. The presence or absence of red blood cell agglutination was determined. The final dilution concentration of the samples that completely inhibited red blood cell agglutination was considered as the HAI antibody titer. This measurement was conducted at an external commercial laboratory routinely performing HAI assays (LSI Medience Corporation, Tokyo, Japan).

Genotyping, quality control, and genome-wide imputation in the Japanese cohorts

The genomic DNA was isolated with the standard protocols from the peripheral blood and genotyped with the use of Infinium Asian Screening Array (Illumina, San Diego, CA, USA). The genotyping probe intensity was converted to SNP genotype calls using Illumina GenomeStudio version 2.0.4 (Illumina, San Diego, CA, USA). We applied stringent QC filters to the genotype data using PLINK2[42] as described elsewhere[51]. Briefly, we excluded samples with a genotyping call rate < 0.98. We included only the samples of the

estimated East Asian ancestry, based on the principal component analysis with the HapMap project samples[40]. We further filtered out SNPs with (i) call rate <0.99; (ii) minor allele count <5; (iii) *P*-value for Hardy-Weinberg equilibrium $<1.0 \times 10^{-10}$; and (iv) with more than 5% allele frequency difference when compared with the representative reference panels of Japanese ancestry (*i.e.*, the reference panel used for the genotype imputation in this study and the allele frequency panel of Tohoku Medical Megabank Project[41]). After QC, we obtained genotype data of 519,668 autosomal and 17,359 X-chromosome SNPs for the 2,145 individuals. Additionally, we performed Sanger sequencing for all the individuals on the three SNPs at the IGH locus, which showed genome-wide significant associations with the SARS-CoV-2-specific antibody titer in our preliminary analysis (see below) using imputed genotypes, incorporating the sequencing-based genotypes into the GWAS data. To extend the coverage of the genetic variants to be tested, we performed genome-wide genotype imputation. We used SHAPEIT4 software[43] version 4.2.1 for haplotype phasing and Minimac4 software[44] version 1.0.1 for genotype imputation. For imputation, we used our in-house and Japanese-specific reference panel composed of n = 11,754 whole-genome sequence (WGS) data from multiple studies (*e.g.,* n = 1,939 from the BBJ study[52] and n = 141 WGS from the previous study[53]).

Preliminary genome-wide association study of the SARS-CoV-2-specific antibody titers

We first performed a GWAS of the SARS-CoV-2-specific antibody titers in the discovery cohort (n = 1,559) measured using the Abbott SARS-CoV-2 IgG II Quantitative Antibody Assay (*i.e.*, assay kit 1). In this preliminary analysis, we performed genotype imputation solely based on the SNPs covered by the SNP array (Infinium Asian Screening Array) and applied a relaxed post-imputation quality control filter of MAF >0.5% and imputation *INFO* >0.3 to the imputed variants. We observed the MHC and IGH loci surpassing the genome-wide significant threshold of $P = 5.0\times10^{-8}$ (Figure S19). The three lead variants at IGH (14:106208306:A:G, 14:106208326:G:C, and 14:106208327:G:A; $P = 3.1\times10^{-11}$, 1.2×10^{-11}, and 1.2×10^{-11}) were imputed variants with low MAF (0.030, 0.028, and 0.028) and moderate imputation *INFO* (0.37, 0.36, and 0.36). To rule out the possibility that these associations were artefacts due to the relatively limited imputation performance, we performed Sanger sequencing on the three SNPs for all the individuals in the discovery and replication cohorts to directly determine the genotypes. We combined these Sanger sequencing-based genotypes of the three SNPs with the SNP array genotype data and again performed genotype imputation for GWAS. The resulting imputed data was used for the final GWASs described in the main text.

Sanger sequencing of rs1043249, rs1043109, and rs193160354

Polymerase chain reactions (PCRs) were carried out using AmpliTaq Gold™ 360 Master Mix (Thermo Fisher Scientific, U.S., Massachusetts) to amplify the target DNA sequences for Sanger sequencing. Primers were designed with the following sequences: forward primer (Fw): GCTGCCCTGTAGGGACAGAGGT and reverse primer (Rv): CGGACCCCTGAGGTCACATG. Each 20 μL reaction mixture consisted of 10 μL of AmpliTaq Gold 360 Master Mix (Thermo Fisher Scientific, U.S., Massachusetts), 0.4 μL of forward primer (10 μM), 0.4 μL of reverse primer (10 μM), 0.5 μL of template DNA, and 8.7 μL of deionized water. The PCR amplification protocol was initiated with a preliminary denaturation step at 95°C for 10 minutes, followed by 35 cycles of denaturation at 95°C for 15 seconds (or 95°C for 10 seconds), annealing at 67°C for 30 seconds, and extension at 72°C for 30 seconds (or 72°C for 1 min). A final extension was performed at 72°C for 7 minutes. The PCR products were then held at 16°C indefinitely. Post-amplification, the PCR products were purified using the ExoSAP-IT™ PCR Product Cleanup Reagent (Thermo Fisher Scientific, U.S., Massachusetts). The purified products were sent to Azenta Japan (Japan, Tokyo) for Sanger sequencing, following the standard protocols provided by the sequencing facility.

UK biobank data

The UKB is composed of health-related information from approximately 500,000 individuals aged between 40 and 69 years recruited from across the United Kingdom from 2006 to 2010. The detailed characteristics of the cohort and genotype–phenotype data were described elsewhere[54]. Briefly, we used the genomic data based on genotyping either by the Applied Biosystems UK BiLEVE Axiom Array or by the Applied Biosystems UK Biobank Axiom Array and imputation using a combination of the Haplotype Reference Consortium, UK10K, and 1KG Phase 3 reference panels. We included only individuals of British ancestry who passed QC. The IgG antibody serostatus (*i.e.*, positive or negative) were derived from the SARS-CoV-2 coronavirus self-test antibody study Phases 1 and 2 (UKB show case Resource 4500 and 4501). In Phase 1, a total of 46,310 participants were recruited between February and April 2021. The participants were sent the Fortress Fast COVID-19 Device kit to self-test the serostatus of anti-S IgG and were asked to report their antibody test results and COVID-19 vaccination status to UKB. In Phase 2, a total of 156,204 participants were recruited between March and May 2021. The participants were sent the AbC-19™ Rapid Test kit to self-test the serostatus of anti-S IgG and were asked to report their antibody test results and COVID-19 vaccination status to UKB. In both Phases, participants who reported positive test results for anti-S IgG were invited to take part in a follow-on study, the SARS-CoV-2 coronavirus infection study, in which participants were sent the Thriva coronavirus antibody test kit and were asked to return their blood sample (UKB show case Resource 4600). The blood sample was tested for anti-N IgG, which is exclusively produced by natural infection but not by vaccination, unlike anti-S IgG. We used anti-N IgG test results to differentiate vaccine-induced immunity and natural immunity.

HLA imputation

We also performed HLA genotype imputation for fine-mapping of the MHC region. We extracted the genotyped SNPs in the extended MHC region (24-36 Mb on chromosome 6, NCBI Build 37). Based on these SNPs, we imputed the classical and non-classical HLA alleles (two-, four-, and six-digits) and corresponding amino acid sequences using DEEP*HLA[45], a multi-task convolutional deep learning method. We used the high-resolution HLA reference panel of the Japanese population[17] (n = 1,118). The HLA imputation procedure produced binary markers indicating the presence or absence of an investigated HLA allele or an amino acid sequence. HLA variants imputed with MAF >0.5% and an imputation quality score (r^2 in cross-validation) >0.7 were used for the subsequent analyses.

Genome-wide association analysis

Japanese vaccination cohorts data

First, we performed genome-wide single-variant association tests of the immunogenicity traits separately for the two Japanese vaccination cohorts. The measured IgG antibody titers or T-cell responses were transformed into a standard normal distribution based on rank before association tests. Associations of the imputed genotype dosages of the genetic variants with each quantitative trait were statistically tested using the whole-genome regression approach implemented in REGENIE[46]. Variants imputed with MAF >0.5% and imputation *INFO* >0.7 were used for the analysis. We incorporated age, sex, age^2, age×sex, age^2×sex, and the top 10 principal components into the regression model as covariates. The resulting association statistics of the two vaccination cohorts were meta-analyzed using the inverse-variance weighted fixed effect model as implemented in METAL[47]. We set a genome-wide significance threshold of $P < 5.0\times10^{-8}$.

UKB data

First, we performed genome-wide single-variant association tests of the binary IgG antibody serostatus (Data-Field 27981) separately for the two self-test kits (*i.e.*, Fortress Fast COVID-19 Test and AbC-19^{TM} Rapid Test). We confined the participants to those who performed the antibody serostatus test at least one week after vaccination. Participants with positive anti-N IgG results in the follow-on study were excluded from the analysis given the potential natural infection. We note that our analysis procedure is different in the inclusion criteria from the previous study analyzing this dataset[3], in which individuals with negative serostatus were limited to those having their antibody test result within 20–60 days after their first vaccine or within 0–300 days after their second vaccine dose. Our analysis procedure aimed to maximize the GWAS sample size to gain statistical power. Consequently, we analyzed 152,906 individuals out of the about 200,000 study participants, whereas the previous study did 54,066 individuals. Associations of the imputed genotype dosages of the genetic variants with the binary serostatus were statistically tested using the whole-genome regression approach implemented in REGENIE. Variants imputed with MAF >0.5% and imputation *INFO* >0.7 were used for the analysis. We incorporated age, sex, age^2, age×sex, age^2×sex, and the top 10 principal components into the regression model as covariates. The resulting association statistics of the two self-test kits were meta-analyzed using the inverse-variance weighted fixed effect model as implemented in METAL[47]. We set a genome-wide significance threshold of $P < 5.0\times10^{-8}$.

PacBio HiFi long-read WGS

DNA samples were sheared targeting the size of 20 kb using Megaruptor 3 (Diagenode). SMRTbell libraries were prepared with the SMRTbell Express Template Prep Kit 2.0 according to the manufacturer's protocols. Fragments were size-selected using SageELF (Sage Science). Libraries were sequenced on the Sequel II (Pacific Bioscience) system using the Sequel II Binding Kit 2.0 and Sequel II Sequencing Kit 2.0. Based on the sequenced subreads, circular consensus sequence (CCS) reads were generated using SMRT Link v9.0.0 (Pacific Bioscience). CCS reads were aligned against hs37d5 using pbmm2 v1.7.0 (https://github.com/PacificBiosciences/pbmm2), and the resulting reads alignments were manually inspected using the Integrative Genomics Viewer[49].

Colocalization analysis

We performed a colocalization analysis of the *IGKV1-13* eQTL and the UKB IgG antibody serostatus associations using COLOC[48] with the default parameters. COLOC estimates whether the association signals of two phenotypes share common causal variants in a given genomic region and computes posterior probabilities for the following five hypotheses: H0, neither trait exhibits a genetic association; H1/H2, only one trait exhibits a genetic association; H3, both traits are associated but with independent causal variants and H4, both traits are associated with a single causal variant.

Association analysis of HLA variants

We statistically tested whether imputed HLA variants were associated with IgG antibody titer or T-cell response using linear regression models as implemented in *R* statistical software version 3.6.3. Accounting for the employment of a linear regression model, we excluded 3rd-degree or more closely related individuals (KING[55] kingship coefficient cutoff >0.0884) from the GWAS dataset. We assumed additive effects of the allelic dosages on the continuous phenotypes. We defined the HLA variants as biallelic single-nucleotide variants in the MHC region, two-, four-, and six-digits biallelic HLA alleles, biallelic HLA amino acid polymorphisms corresponding to their respective residues, and multiallelic HLA amino acid polymorphisms for each amino acid position. We incorporated cohort, age, sex, age^2, age×sex, age^2×sex, and the top 10 principal components into the regression model. For multiallelic amino acid variants, we estimated its significance by an omnibus test for each amino acid position by a log-likelihood ratio test, comparing

the likelihood of the fitted model with the null model. The significance of the improvement of the model fitting was evaluated by the deviance, which follows χ^2 distribution with $m - 1$ degree(s) of freedom for an amino acid position with m polymorphic residues. The conditional association analysis was performed to find additional HLA genes associated with the continuous phenotypes by additionally including the HLA allele/haplotype as covariates in the regression model. We tested a multivariate full regression model by incorporating the associated amino acid polymorphisms and classical HLA alleles, with the most frequent residue excluded from each amino acid position as the reference allele. The proportion of variance explained by each variant was computed using the anova() function implemented in *R* statistical software.

Protein QTL analysis

To functionally characterize the vaccine immunogenicity-associated variants, we analyzed the proteogenomic datasets from two Japanese cohorts: Biobank Japan (BBJ) and Japan COVID-19 Task Force (JCTF). The pQTL analysis was performed separately for the two cohorts as described below and then were meta-analyzed using the inverse-variance weighted fixed effect model.

Biobank Japan

BBJ is a multi-institutional hospital-based registry that collaboratively recruited approximately 260,000 patients with at least one of 51 target diseases and collected DNA, serum samples, and medical records from individuals of Japanese ancestry[19]. Blood-derived DNA samples were genotyped using the Illumina HumanOmniExpressExome BeadChip or a combination of the Illumina HumanOmniExpress and HumanExome BeadChips. We excluded samples with a genotyping call rate <0.98. We included only the samples of the estimated East Asian ancestry, based on the principal component analysis with the HapMap project samples[40]. We applied standard QC filters to variants, excluding those with (i) SNP call rate <0.99, (ii) minor allele count <5, and (iii) Hardy–Weinberg equilibrium P value $<1.0 \times 10^{-6}$. We used SHAPEIT4 software[43] version 4.2.1 for haplotype phasing and Minimac4 software[44] version 1.0.1 for genotype imputation with our in-house reference panel described above (n = 11,754). The protein concentrations were measured in the serum samples from 2,700 individuals using the Olink Explore 3072 platform across three batches. The resulting protein expression levels in a normalized scale (Normalized Protein eXpression; NPX) were bridge-normalized using OlinkAnalyze R package. The bridge-normalized NPX values were transformed into a standard normal distribution based on rank and then adjusted for age, sex, age^2, age×sex, age^2×sex, disease status for the 47 target diseases in the BBJ, top 20 PCs of the transformed NPX matrix, top 10 PCs of genotypes, and protein measurement batch in a linear regression model. We then transformed the resulting residuals into a standard normal distribution on rank. For the vaccine immunogenicity-associated variants within MHC, given the multiple HLA variant associations (*i.e.*, amino acid polymorphisms and classical HLA alleles), we calculated HLA genetic scores (HLA-GSs) for BBJ individuals based on the effect sizes of the associated HLA variants estimated by the multivariate regression models in the HLA fine-mapping analysis. The associations of the derived HLA-GSs with the protein expressions were evaluated by linear regression models. The association tests for variants outside MHC were performed using the whole-genome regression approach implemented in REGENIE.

Japan COVID-19 Task Force

The Japan COVID-19 Task Force (JCTF) was established in early 2020 as a nationwide multicenter consortium and collected DNA, plasma samples, and medical records from patients diagnosed with COVID-19[7]. The COVID-19 severity was categorized into four levels: "Most severe" for patients in ICU or requiring intubation and ventilation, Severe for others requiring oxygen support, Mild for other symptomatic patients (*e.g.* shortness of breath), and "Asymptomatic" for those without COVID-19-related symptoms. Blood-derived DNA samples were genotyped using the Infinium Asian Screening Array. We excluded samples with a genotyping call rate <0.98. We included only the samples of the estimated East Asian ancestry, based on the principal component analysis with the HapMap project samples[40]. We applied standard QC filters to variants, excluding those with (i) SNP call rate <0.99, (ii) minor allele count <5, and (iii) Hardy–Weinberg equilibrium P value $<1.0 \times 10^{-10}$. We used SHAPEIT4 software[43] version 4.2.1 for haplotype phasing and Minimac4 software[44] version 1.0.1 for genotype imputation with our in-house reference panel described above (n = 11,754). The protein concentrations were measured in the plasma samples from 1,536 individuals using the Olink Explore 3072 platform across three batches. The resulting NPX were bridge-normalized using OlinkAnalyze R package. The bridge-normalized NPX values were transformed into a standard normal distribution based on rank and then adjusted for age, sex, age^2, age×sex, age^2×sex, COVID-19 severity, top 20 PCs of the transformed NPX matrix, top 10 PCs of genotypes, and protein measurement batch in a linear regression model. We then transformed the resulting residuals into a standard normal distribution on rank. For the vaccine immunogenicity-associated variants within MHC, we calculated HLA-GSs for JCTF individuals based on the effect sizes of the associated HLA variants estimated by the multivariate regression models in the HLA fine-mapping analysis. The associations of the derived HLA-GSs with the protein expressions were evaluated by linear regression models. The association tests for variants outside MHC were performed using the whole-genome regression approach implemented in REGENIE.

Hematopoietic mosaic chromosomal alteration in the UKB participants

mCA calls generated in the previous study[25] from UK Biobank participants were used in our analysis. The mCA call data were obtained from dataset Return 3094 generated from UKB application 47821. This dataset contained 2,388 gain, 3,718 loss, 8,184 copy-neutral loss of heterozygosity (CN-LOH), and 5,339 undetermined copy number change autosomal mCA events detected among 482,744 participants from UKB. Of these, 791 gain, 2,328 loss, and 1,235 CN-LOH mCA events were found with >10% cell fraction, defined as expanded mCAs. This data also contains mLOX and mLOY calls. To focus on high-confidence mLOX/mLOY events, we

applied strict filtering to the call set. mLOX events were restricted to those annotated as copy number loss and exceeding 100Mbp in size. mLOY events were restricted to those annotated as copy number loss and exceeding 2Mbp in size. The resulting data contained 825 expanded mLOX with cell fraction >5%, 410 expanded mLOX with cell fraction >10%, 17,947 expanded mLOY with cell fraction >5%, and 9,813 expanded mLOY with cell fraction >10%.

Association testing between vaccine-induced antibody serostatus and mCAs

To identify classes of mCAs affecting the vaccine-induced serostatus, mCA events were classified according to chromosomal location and copy number shift (*i.e.,* gain, loss, or copy-neutral loss-of-heterozygosity [CN-LOH]). Loss and CN-LOH events were further subdivided by p-arm and q-arm. The associations between mCA classes and the serostatus were evaluated using a logistic regression model adjusted for adjusted for age, age^2, sex, antibody self-test kit, number of doses, dose intervals, smoking status, Townsend deprivation, BMI, and the top 10 PCs. We excluded mCA events with less than 30 carriers from the analysis considering the limited statistical power as in earlier work[25]. The robustness of the mCA association with the serostatus was further evaluated by performing follow-up association testing while excluding individuals having potential immunosuppression-related traits, namely cancers, aplastic anemia, neutropenia, chemotherapy, bone marrow or stem cell transplant, and radiotherapy. The correspondence table of the excluded traits and UKB Data-Field is shown in Table S7.

Hematopoietic mosaic chromosomal alteration in the biobank Japan participants

To investigate the effect of mCAs at MHC or IGH loci on infectious/immune diseases, we analyzed the Biobank Japan data. We calculated B-allele frequency (relative allelic intensity) and $\log_2 R$ ratio (total allelic intensity) values using the intensity data derived from the genotyping arrays and integrated them with the phased genotypes described above. Mosaic chromosomal alteration detection was performed using the MoChA pipeline[24,25] (https://github.com/freeseek/mocha/). In addition to the default mCA call filters indicated in the pipeline documentation, we further excluded mCA events with (i) relative coverage <0.6 or (ii) length <2 Mbp and relative coverage >2.5 to filter out potential constitutional deletions or duplications. The resulting call set contained 3,973 gain, 4,417 loss, 9,521 CN-LOH, and 7,888 undetermined copy number change autosomal mCA events detected among 180,389 participants from BBJ. Of these, 615 gain, 1,914 loss, and 966 CN-LOH, and 438 undetermined copy number change mCA events were found with >10% cell fraction, defined as expanded mCAs. The MoChA pipeline also detects mLOX/mLOY events. mLOX events were restricted to those annotated as copy number loss and exceeding 100Mbp in size. mLOY events were restricted to those annotated as copy number loss and exceeding 2Mbp in size. The resulting data contained 331 expanded mLOX with cell fraction >5%, 74 expanded mLOX with cell fraction >10%, 9,231 expanded mLOY with cell fraction >5%, and 6,960 expanded mLOY with cell fraction >10%.

Association testing between infectious/immune diseases and mCAs

MHC-affecting and IGH-affecting mCA events were defined as those overlapping with the genomic regions of 6:28.5–33.4 Mbp and 14:106.1–107.2 Mbp (GRCh37), respectively. The associations between infectious/immune diseases and MHC/IGH-affecting mCAs were evaluated using a logistic regression model adjusted for age, age^2, sex, smoking status, and the top 10 PCs. The infectious/immune disease statuses for BBJ participants were derived from past medical history records and follow-up survey records retrieved from electronic medical records, and those for UKB participants were derived from ICD-10-coded diagnoses records (Data-Field 41270). The correspondence table of the evaluated infectious/immune diseases and ICD-10 codes is shown in Table S8. In this analysis, we tested all the phenotypes with at least two cases having the mCA events of interest to provide the as comprehensive landscape of these mCA associations with infectious or immune diseases as possible. We tested if known germline Graves' disease risk HLA alleles[56] (HLA-DPβ1 Leu35 and His9; HLA-A Phe9, Tyr9, and Thr9; HLA-B Glu45, Phe67, Tyr67, and Cys67; and HLA-DRβ1 Glu74) were associated with the acquisition of MHC-affecting mCA in BBJ, confirming no confounding effect.