

IgG4-related disease in the Japanese population: a whole-genome sequencing study

*Yuxun Oswald Zhang, Takeshi Iwasaki, Takahisa Kawaguchi, Hiroki Takahashi, Shuji Kawaguchi, Atsushi Kanno, Izumi Yamaguchi, Kensuke Kubota, Hiroaki Dobashi, Masao Nagasaki, Motohisa Yamamoto, Meiko Takahashi, Masakazu Shimizu, Tsukasa Ikeura, Shoko Matsui, Masatoshi Kanda, Koki Nakamura, Kensuke Yokoyama, Atsushi Azumi, Yasufumi Masaki, Ichiro Mizushima, Yusuke Kurita, Hiroshi Seno, Tomoki Origuchi, Shujiro Yazumi, Kenji Hirano, Atsushi Masamune, Nobumasa Mizuno, Hiromi Shimada, Masafumi Moriyama, Yasuki Hori, Yuzo Kodama, Takako Saeki, Toshifumi Kin, Chiharu Kawanami, Masanori Asada, Takashi Akamizu, Akira Nakamura, Koichi Oshima, Yoshiya Tanaka, Hajime Yoshifuji, Terumi Kamisawa, Toshiyuki Kimura, Hisanori Umehara, Hideki Ishikawa, Tsutomu Chiba, Kazuichi Okazaki, Tsuneyo Mimori, Seiji Nakamura, Mitsuhiro Kawano, Fumihiko Matsuda, on behalf of the Japanese IgG4-Related Disease Working Consortium**

*Consortium members are listed at the end of the Article.

Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan (Y O Zhang MPH, T Iwasaki MD PhD, T Kawaguchi PhD, S Kawaguchi PhD, I Yamaguchi PhD, Prof M Nagasaki PhD, M Takahashi PhD, M Shimizu PhD, Prof F Matsuda PhD); **Department of Rheumatology and Clinical Immunology, Sapporo Medical University School of Medicine, Sapporo, Japan** (Prof H Takahashi MD PhD, M Kanda MD PhD, K Nakamura MD); **Department of Medicine, Division of Gastroenterology, Jichi Medical University, Shimotsuke, Japan** (A Kanno MD PhD, K Yokoyama MD PhD); **Department of Oncology, Yokohama City University Graduate School of Medicine, Yokohama, Japan** (Prof K Kubota MD PhD, Y Kurita MD PhD); **Division of Hematology, Rheumatology and Respiratory Medicine, Department of Internal Medicine, Faculty of Medicine, Kagawa University, Kagawa, Japan** (H Dobashi MD PhD, H Shimada MD PhD); **Department of Rheumatology and Allergy, IMSUT Hospital, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan** (M Yamamoto MD PhD);

Department of Gastroenterology and Hepatology, Kansai Medical University, Hirakata, Japan (T Ikeura MD PhD); First Department of Internal Medicine, University of Toyama, Toyama, Japan (Prof S Matsui MD PhD); Department of Ophthalmology, Kobe Kaisei Hospital, Kobe, Japan (A Azumi MD PhD); Department of Hematology and Immunology, Kanazawa Medical University, Uchinada, Japan (Prof Y Masaki MD PhD, Prof M Kawano); Department of Nephrology and Rheumatology, Kanazawa University, Kanazawa, Japan (I Mizushima MD PhD); Department of Gastroenterology and Hepatology, Kyoto University Graduate School of Medicine, Kyoto, Japan (Prof H Seno MD PhD, Prof T Chiba MD PhD); Department of Immunology and Rheumatology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan (Prof T Origuchi MD PhD); Department of Gastroenterology and Hepatology, Medical Research Institute, Kitano Hospital, PIIF Tazuke-Kofukai, Osaka, Japan (S Yazumi MD PhD); Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan (K Hirano MD PhD); Department of Gastroenterology, Tohoku University School of Medicine, Sendai, Japan (Prof A Masamune MD PhD); Department of Gastroenterology, Aichi Cancer Center Hospital, Nagoya, Japan (N Mizuno MD PhD); Section of Oral and Maxillofacial Oncology, Division of Maxillofacial Diagnostic and Surgical Sciences, Faculty of Dental Science, Kyushu University, Fukuoka, Japan (Prof M Moriyama DDS PhD, Prof S Nakamura DDS PhD); Department of Gastroenterology and Metabolism, Graduate School of Medical Sciences, Nagoya City University, Nagoya, Japan (Y Hori MD PhD); Department of Gastroenterology, Kobe University Graduate School of Medicine, Kobe, Japan (Y Kodama MD PhD); Department of Internal Medicine, Nagaoka Red Cross Hospital, Nagaoka, Japan (T Saeki MD PhD); Center for Gastroenterology, Teine-Keijinkai Hospital, Sapporo, Japan (T Kin MD); Department of Gastroenterology, Japanese Red Cross Otsu Hospital, Otsu, Japan (C Kawanami MD); Department of Gastroenterology, Osaka Red Cross Hospital, Osaka, Japan (M Asada MD PhD); First Department of Medicine, Wakayama Medical University, Wakayama, Japan and Center for Excellence in Thyroid Care, Kuma Hospital, Kobe, Japan (Prof T Akamizu MD PhD); Second Department of Internal Medicine, Shinshu University School of Medicine, Matsumoto, Japan (A Nakamura MD PhD);

Department of Ophthalmology, National Hospital Organization, Okayama Medical. Center, Okayama, Japan (K Oshima MD); The First Department of Internal Medicine, University of Occupational and Environmental Health Japan, Kitakyushu, Japan (Prof Y Tanaka MD PhD); Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan (H Yoshifuji MD PhD, Prof T Mimori MD PhD); Tokyo Metropolitan Cancer and Infectious Diseases Center Komagome Hospital, Tokyo, Japan (T Kamisawa MD PhD); Department of Gastroenterology, Hyogo Prefectural Amagasaki General Medical Center, Amagasaki, Japan (T Kimura MD); Center for RA and Autoimmune Diseases, Nagahama City Hospital, Nagahama, Japan (Prof H Umehara MD PhD); Department of Molecular-Targeting Cancer Prevention, Kyoto Prefectural University of Medicine, Kyoto, Japan (H Ishikawa MD PhD); Department of Internal Medicine, Kansai Medical University Kori Hospital, Neyagawa, Japan (Prof K Okazaki MD PhD)

Correspondence to:

Prof Fumihiko Matsuda, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto 606–8507, Japan

fumi@genome.med.kyoto-u.ac.jp

Summary

Background

IgG4-related disease is a rare autoimmune disorder characterised by tissue infiltration of IgG4-positive plasma cells, storiform fibrosis, elevated serum IgG4 concentrations, and increased risk of tumour complications. Previous genetic studies have implicated *FCGR2B* and *HLA* loci in susceptibility to IgG4-related disease; however, most relied on microarray-based genotyping and imputation, which have limited resolution in highly polymorphic and structurally complex regions. This study aimed to investigate genetic susceptibility to IgG4-related disease using comprehensive genomic variant analysis, including low-frequency and structural variants not readily captured by microarrays.

Methods

We conducted a genome-wide association study using whole-genome sequencing data in a two-set, subtype-stratified case–control design in the Japanese population. Set 1 included samples (cases) from patients with IgG4-related disease from 50 hospitals across Japan participating in the Japanese IgG4-Related disease Working Consortium (recruited between Oct 27, 2008, and March 3, 2016) and previously sequenced Japanese population control samples. Set 2 included samples (cases) from patients with IgG4-related disease from eight hospitals of the Consortium (recruited between Aug 12, 2021, and Dec 20, 2023) and previously sequenced healthy individuals residing in the Tokyo metropolitan area. No specific inclusion or exclusion criteria were applied to either cases and controls. We used whole-genome sequencing at depths of 15× or 30× using HiSeqX and NovaSeq platforms (Illumina; San Diego, CA, USA) to enable the inclusion of previously uncaptured single-nucleotide polymorphisms and direct analysis of HLA amino acid residues. We investigated complement component 4 copy number variations using short-read sequencing data and established read-depth-based typing methods. People with lived experience of IgG4-related disease were not involved in the study.

Findings

This whole-genome sequencing study comprised of 2 sets. Set 1 included 646 patient samples (172 [26.6%] were female, 474 [73.4%] were male, and the mean age was 64.4 years [SD 11.4]) and 2254 population controls (1348 [59.8%] were female and 906 [40.2%] were male). Set 2 included 223 patient samples (78 [35.0%] were female, 144 [65.0%] were male, and the mean age was 63.5 years [10.9]) and 405 population controls (65 [16.0%] were female and 340 [84.0%] were male). All individuals were of Hondo Japanese ancestry. The average IgG4 concentration at diagnosis was 653.1 mg/dL (SD 596.3) in Set 1 and 543.5 mg/dL (603.5) in Set 2. We validated the *FCGR2B* ($p=9.8\times10^{-11}$) region as the susceptibility locus for IgG4-related disease. *PTCH1* ($p=3.8\times10^{-8}$) and long non-coding RNA *LOC102724227* were found to be specific susceptibility loci for Mikulicz's disease. We also confirmed the association between the HLA amino acid residue DRB1-GB-7 ($p=1.1\times10^{-19}$) with IgG4-related disease and identified two additional residues, A-GA2-9 ($p=4.1\times10^{-6}$) and DQB1-GB-82 ($p=4.7\times10^{-9}$), that were significantly associated with IgG4-related disease. In the joint-association analysis of complement component 4 copy number variation, *C4A* showed a protective association with IgG4-related disease ($\beta=-0.127$, $p=7.9\times10^{-3}$), whereas *C4B* was associated with an increased risk ($\beta=0.151$, $p=1.9\times10^{-2}$). A low level of linkage disequilibrium ($r^2<0.15$) was observed between the *C4A* and *C4B* alleles and the identified HLA amino acid residues in the main island Japanese population.

Interpretation

C4 copy number variation, in addition to *HLA* and *FCGR2B*, was found to be a distinct genetic factor associated with IgG4-related disease susceptibility, illustrating the complex polygenic nature of the disease. Furthermore, the identification of *PTCH1* and the long non-coding RNA *LOC102724227* as Mikulicz's disease-specific susceptibility loci suggests that genetic heterogeneity might underlie the clinical diversity of IgG4-related disease, particularly with respect to the affected organs.

Funding

The Japanese Ministry of Health, Labour, and Welfare, the Japanese Agency of Medical Research and Development, the Kyoto University Grant for Top Global University Japan Project, and the Kyoto University Division of Graduate Studies SPRING Program.

Introduction

IgG4-related disease is a rare autoimmune disorder characterised by tissue infiltration of IgG4-positive plasma cells, storiform fibrosis, elevated serum IgG4 concentrations, and an elevated incidence of tumour complications.¹ IgG4-related disease was first recognised in 2001.² IgG4-related disease encompasses various organ-specific conditions with shared histopathological features, such as IgG4-related autoimmune pancreatitis, IgG4-related dacryoadenitis and sialadenitis (collectively known as Mikulicz's disease), Riedel's thyroiditis, and retroperitoneal fibrosis.³ The annual incidence of IgG4-related disease in Japan was reported to range from 0.28 per 100 000 individuals to 1.08 per 100 000 individuals in 2009.⁴ However, this incidence has been increasing annually. A 2016 nationwide epidemiological survey focusing on autoimmune pancreatitis estimated annual incidence at 3.01 per 100 000 individuals.⁵ This substantial increase is attributed to greater awareness of the disease concept and refinement of the diagnostic criteria. In the USA, a claims-based algorithm estimated the annual incidence of IgG4-related disease to range from 0.78 per 100 000 to 1.39 per 100 000 between 2015 and 2019.⁶ Although most reported cases (approximately 75%) are Japanese, IgG4-related disease has been observed across nearly all ethnic groups.⁶

Genetic predisposition is a known risk of IgG4-related disease. Previous studies have highlighted the contribution of human leukocyte antigen (*HLA*) and non-*HLA* gene variants to IgG4-related disease susceptibility. In a previous genome-wide association study (GWAS) using microarray data, the *FCGR2B* and *HLA-DRB1* regions were identified as susceptibility loci, with an additional amino acid residue-level association at the seventh amino acid residue within the β domain of the HLA-DRB1 peptide-binding groove.⁷ Although the GWAS was performed using single nucleotide variant arrays, the analysis also included imputed variants. However, the variants used for the association analysis were restricted to those with a minor allele frequency of 1% or greater. Consequently, rare variants could not be evaluated. Furthermore, single nucleotide variant arrays are inherently unsuitable for the analysis of structural variants and copy number variations. The complement component 4 (*C4*) gene exemplifies this

limitation, because emerging evidence suggests that copy number variations within this gene are associated with an increased risk of systemic inflammatory autoimmune diseases.⁸ Thus, comprehensive evaluation of *C4* copy number variations is required to elucidate the genetic architecture underlying IgG4-related disease. Additionally, the diverse organ involvement in IgG4-related disease subtypes suggests distinct genetic mechanisms that stratified analysis could help clarify regarding pathophysiological differences.

To address these limitations, we conducted whole-genome sequencing in a Japanese cohort to explore the genetic landscape of IgG4-related disease, including stratified analyses by disease subtypes. Using short-read whole-genome sequencing, we performed a GWAS incorporating rare single nucleotide variants and variants undetectable by the microarray methods while enabling comprehensive *HLA* allele calling across Class I, Class II, and Class III regions previously restricted by imputation panel resources.⁷ Using sequencing data and established typing methods, we further investigated the association between *C4* copy number variations and IgG4-related disease. Importantly, all analyses were performed without genotype imputation.

Methods

Study design and participants

We conducted a two-set, subtype-stratified case–control study to perform a GWAS with WGS data investigating IgG4-related disease susceptibility. Between Oct 27, 2008, and March 3, 2016, we recruited patients who fulfilled the international comprehensive diagnostic criteria for IgG4-related disease from 50 departments belonging to 48 institutes participating in the Japanese IgG4-Related Disease Working Consortium as part of our previous GWAS study and designated as cases in Set 1 (appendix pp 20–21).^{7,9} Whole-genome sequencing data of Japanese individuals previously sequenced were used as controls for Set 1.¹⁰ No specific inclusion or exclusion criteria were applied to either cases or controls.

For Set 2, between Aug 12, 2021, and Dec 20, 2023, we recruited patients who fulfilled the same criteria from eight institutes belonging to the Consortium (appendix pp 20–21). The control group for Set 2 comprised healthy individuals residing in the Tokyo metropolitan area whose genomes had been previously sequenced.¹¹ Again, no specific inclusion or exclusion criteria were applied. There was no overlap of individuals between case or control samples in Set 1 and Set 2.

The diagnosis of IgG4-related disease subtypes, autoimmune pancreatitis and Mikulicz's disease, was made in accordance with the international comprehensive diagnostic criteria.¹² For Set 1, a single physician (IY) made the diagnoses based on clinical information provided by each patient's attending physician. In contrast, for Set 2, the diagnoses were made directly by the attending physicians themselves. For subtype analyses, patients with single or coexisting subtypes were classified into each respective subtype.

This study complied with the Declaration of Helsinki and was approved by the Ethics Committees of Kyoto University (G0439-7 and R1892-14). All patients were adequately informed about the purpose and procedures of this study, and written consent was obtained

from each participant. People with lived experience of IgG4-related disease were not involved in the design, analysis, or interpretation of the study. All analyses were based on clinical and genetic data, without direct engagement of participants representatives in shaping research questions, selecting outcome measures, or disseminating findings.

Whole genome sequencing, variant calling, and quality control

In Set 1, patients underwent whole-genome sequencing using the NovaSeq platform (Illumina; San Diego, CA, USA): patients had sequencing to a mean depth of 15 or 30, depending on instrument performance at time of sequencing. The control group was sequenced using the HiSeqX platform (Illumina; San Diego, CA, USA) to a depth of 15×. In Set 2, all patients were sequenced using the NovaSeq platform (Illumina; San Diego, CA, USA) to a mean depth of 30, and the control group were sequenced using the HiSeqX platform to a mean depth of 30 (appendix p 22).

The whole-genome sequencing results from cases and controls used in this study were aligned and called in a single run across samples in this study and other genomic studies.¹³ We conducted quality control of the sequence data by Genome Analysis Toolkit (version 4.2.6.1) Best Practices to remove variants with low quality, heterozygous genotypes with imbalanced allele ratios, and genotypes with insufficient confidence or coverage, with thresholds defined as variant quality score log-odds <10, mapping quality <58-75, depth of coverage <7, and autosomal heterozygosity rates outside ± 5 standard deviations of the mean (appendix pp 4, 6). Population genetics studies show that the Japanese population can be divided into two major clusters: the Hondo cluster (main islands excluding Okinawa), comprising approximately 98% of the population, and the Ryukyu cluster (Okinawa and surrounding islands), comprising about 2% of the population. These clusters differ substantially at the genomic level.¹⁰ Unequal representation of the Ryukyu cluster between cases and controls might lead to false-positive associations unrelated to disease. Therefore, this analysis was restricted to individuals from the Hondo cluster. To select cases and controls belonging to the Hondo cluster, we conducted

principal component analysis using genome-scan results of 183 042 individuals from the BioBank Japan dataset.¹⁴ 84 483 single nucleotide variants with a minor allele frequency of 1% or more that were shared among BioBank Japan and the Set 1 and Set 2 datasets were selected (appendix p 7 for details of the single nucleotide variant selection process) except those within the major histocompatibility complex (*MHC*) locus. Set 1 and Set 2 datasets were projected onto the first and second principal components of the BioBank Japan datasets, and individuals clustering with the Hondo cluster were selected for the association analysis (appendix p 8).

For the quality control steps, all procedures were performed using PLINK (version 2.0).¹⁵ First, we excluded samples with low call rates (below 96% for Set 1 and below 98% for Set 2), heterozygosity rates outside the mean plus or minus 5 SDs, or cryptic relatedness (pairwise identity-by-descent $PI_HAT > 0.2$ using the PLINK identity-by-descent estimator). Subsequently, we performed quality control of the single nucleotide variants using data from the remaining individuals who passed the above criteria. Single nucleotide variants that met any of the following criteria in cases or controls were excluded: a call rate below 98% or a Hardy–Weinberg equilibrium p value below 1.0×10^{-6} . Single nucleotide variants with an extremely low allele count in the merged case and control populations were also excluded. The remaining single-nucleotide variants in each set was used in that set’s GWAS, and the intersection across sets was used for the meta-analysis. For the calculation of principal components used as covariates in the GWAS of Set 1 and Set 2, the following single nucleotide variants were excluded from the variants used for association testing: variants with a minor allele frequency below 1%, all short insertions and deletions, and variants with an Hardy–Weinberg equilibrium p value below 0.05.

Genome-wide association study

For both Sets 1 and 2, we performed a GWAS using Wald’s tests (implemented in RVTESTS, version 20171009).¹⁶ The top ten principal components derived from the principal component analysis and genetically inferred sex were included as covariates. We assessed study statistics

inflation using the genomic inflation factor and visualised through quantile–quantile plots. The genome-wide significance threshold was set at $p=5\cdot0\times10^{-8}$. Association results from both sets were combined using inverse-variance meta-analysis, incorporating effect size estimates, p values, effect allele frequencies, and sample sizes via METAL software (released March 25, 2011).¹⁷

Post-GWAS analysis

For single nucleotide variants identified through GWAS, we applied FINEMAP (version 1.4.1)¹⁸ to refine the association signals, assuming a single causal single nucleotide variant within a 2-megabase region centred on each lead single nucleotide variant to establish the credible set. For the intergenic single nucleotide variants identified in the Mikulicz’s disease cohort, we estimated the variant effect on the neighbourhoods’ gene expression using MENTR (released March 27, 2023), which combines multiple machine-learning models to predict regulatory effects of noncoding variants.¹⁹ We evaluated the association of variants within the credible set from the fine-mapping analysis, incorporating non-coding RNA transcription within a plus or minus 100 kb window around each single nucleotide variant. We adopted the thresholds to 0.1 (robust) and 0.05 (permissive), as proposed by MENTR.

HLA variant identification and association

For *HLA* variant analysis, we selected samples from Set 1 and Set 2 that passed quality control criteria. Using whole-genome sequencing data, *HLA* alleles were determined using the HLA-HD software (version 1.5.0).²⁰ We typed 13 functional Class I and Class II *HLA* genes: *HLA-A*, *HLA-B*, *HLA-C*, *HLA-E*, *HLA-G*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DMA*, *HLA-DMB*, *HLA-DQB1*, *HLA-DRB1*, *HLA-DRB3*, and *HLA-DRB4* genes. The *HLA* dictionary was used to map *HLA* alleles to their respective exon positions and amino acids, and the frequency of each alternative and reference amino acid at each position was computed. To assess the association between amino acid residues and disease susceptibility, we applied a logistic regression model (Python stats model package, version 0.14.0), considering the additive effects of *HLA* amino acid

residue dosages (omnibus test).²¹ Subsequently, we performed iterative conditioning, namely, adding the predominant amino acid variant as a covariate in the regression model, to pinpoint independent candidates. *HLA* amino acid residue positions were determined following the unique renumbering system established by the IPD-IMGT/HLA database.²² Protein structures were visualised using data from the RCSB Protein Data Bank and processed with the UCSF Chimera.²³ For statistical validation of *HLA* amino acid residues, we applied the Bonferroni correction to control the family-wise error rate across multiple tests ($p=2.2 \times 10^{-4}$).

Estimation of *C4* copy number

The *C4* gene has two isotypes: *C4A* and *C4B*. The nucleotide sequences of *C4A* and *C4B* differ at five positions, four of which are non-synonymous substitutions, resulting in four amino acid differences in their respective protein sequences.²⁴ For *C4* copy number variation analysis, we selected DNA samples that met quality control criteria from Set 1 and Set 2. We estimated copy numbers using the GermlineCNVCaller tool in the Genome Analysis Toolkit (version 4.2.6.1), using read-depth information from whole-genome sequencing data.^{25,26} Additionally, the copy numbers of the *C4* isoforms (*C4A* and *C4B*) were estimated using the Genome Analysis Toolkit HaplotypeCaller, based on the read depth of sequences encompassing the five distinguishing nucleotide variants. Detailed methods are in the appendix (pp 4–5, 9).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

After principal component analysis to select individuals belonging to the Hondo cluster and subsequent quality control steps, Set 1 included 646 patient samples (172 [26.6%] were female, 474 [73.4%] were male, and the mean age was 64.4 years [SD 11.4]) and 2254 population controls (1348 [59.8%] were female and 906 [40.2%] were male). 196 (30.3%) of

646 patients with IgG4-related disease had Mikulicz's disease only, 322 (49.8%) had autoimmune pancreatitis only, 72 (11.1%) had both subtypes, and 56 (8.7%) were without subtype information. Set 2 included 223 patient samples (78 [35.0%] were female, 144 [65.0%] were male, and the mean age was 63.5 years [SD 10.9]) and 405 population controls (65 [16.0%] were female and 340 [84.0%] were male). 65 (29.1%) patients had Mikulicz's disease only, 33 (14.8%) had autoimmune pancreatitis only, 30 (13.5%) had both phenotypes, and 95 (42.6%) were without subtype information. The average IgG4 concentration at diagnosis was 653.1 mg/dL (SD 596.3) in Set 1 and 543.5 mg/dL (603.5) in Set 2.

Association analysis was conducted on 5 031 998 single nucleotide variants in Set 1 and 9 282 360 single nucleotide variants in Set 2, both identified by whole-genome sequencing study. The meta-analysis combining Sets 1 and 2 incorporated 4 847 583 single nucleotide variants shared between the two datasets, yielding genomic inflation factor values of 1.028 for the IgG4-related disease cohort, 1.005 for the Mikulicz's disease cohort, and 1.022 for the autoimmune pancreatitis cohort.

In the meta-analysis of the IgG4-related disease cohort, the strongest association peak appeared in the *MHC* locus on chromosome 6. However, the extreme polymorphism of this region led to poor mapping quality and substantial missing data, especially in *HLA* exons. To overcome this issue, we inferred *HLA* alleles using HLA-HD and analysed amino acid residues in the antigen-presenting domains (see later). Four statistically significant single nucleotide variants were identified near the *FCGR2B* locus outside the *MHC* locus. The lead single nucleotide variant, rs13376485, located within the intronic region of *FCGR2B*, was not included in the microarray panel used in a previously published study (figure 1).⁷ The alternative allele of this single nucleotide variant was associated with a decreased disease risk, with an odds ratio (OR) of 0.59 (95% CI 0.51–0.69; $p=9.8\times10^{-11}$; table 2; appendix p 23). The allele frequency was lower in cases (0.147) than in controls (0.232).

We previously identified the protective allele rs1340976,⁷ which showed protective trend (OR 0.72 [95% CI 0.64–0.81]), but not statistically significant ($p=5.7\times10^{-8}$) in this study and displayed low linkage disequilibrium ($r^2=0.18$) with rs13376485 in the pooled dataset. The linkage disequilibrium levels between rs1340976 and the newly detected this single nucleotide variants are in the appendix (p 10). After conditioning on rs13376485, no other single nucleotide variants in this region remained significant; the regional association plot of *FCGR2B* is shown in the appendix (p 11). Fine-mapping analysis identified rs13376485 as the sole variant in the 95% credible set for the target region, with a posterior probability of 0.978 (appendix p 24).

In the meta-analysis of the Mikulicz's disease cohort, an intergenic variant on chromosome 9 near the *PTCH1* gene and a variant on chromosome 12 downstream of the long non-coding RNA *LOC102724227* were identified. rs66917126 in the *PTCH1* region was associated with an elevated disease risk, with the alternative T nucleotide showing an OR of 1.94 (95% CI 1.53–2.46; $p=3.8\times10^{-8}$; appendix p 23). rs66917126 was significantly associated with Mikulicz's disease even in Set 1 alone, with an OR of 2.17 (95% CI 1.66–2.82; $p=1.1\times10^{-8}$; appendix pp 12, 23). No additional associations remained after conditioning on rs66917126 (appendix p 11). Fine-mapping analysis showed that rs66917126 had the highest posterior inclusion probability of 0.495 among 3676 single nucleotide variants (appendix p 24). Three additional single nucleotide variants, rs72754333 and rs72754331 and rs28675274 did not reach the significance threshold in the GWAS ($p\geq5.0\times10^{-8}$); however, their posterior inclusion probabilities were greater than 0.05. For the variant rs894063168 in the *LOC102724227* region, the alternative allele (T) was associated with an increased disease risk, with an OR of 12.57 (5.11–30.93; $p=3.6\times10^{-8}$; appendix p 23). No additional associations remained after conditioning on rs894063168 (appendix p 11). Fine-mapping results showed that rs894063168 had the highest posterior inclusion probability of 0.374 among 3602 single nucleotide variants (appendix p 24). Four additional SNVs, namely, rs1324549138, rs1231784939, an insertion of AATA at chr12:19,203,824, and rs149235894 had a posterior inclusion probability greater than

0.05. Among the single-nucleotide variants identified in the meta-analysis, rs66917126 showed significant heterogeneity of effect sizes between IgG4-related disease and Mikulicz's disease ($Q=7.058$; $p=0.0079$) and between Mikulicz's disease and autoimmune pancreatitis ($Q=9.463$; $p=0.0021$; appendix p 25).

A publicly available expression quantitative trait loci (eQTL) results in the *PTCH1* region showed that both rs66917126 and 15 linkage disequilibrium variants ($r^2>0.2$) were associated with *PTCH1* expression (appendix p 26).²⁷ Neither rs894063168 nor the single-nucleotide variants in linkage disequilibrium with it were associated with *LOC102724227* expression. MENTR predicted mutation effects for rs66917126, rs72754333, and rs34299394 at the *PTCH1* promoter in olfactory epithelial cells (-0.052 effect size), at two enhancers, ADDG09098315655.E in the alimentary part of gastrointestinal system (0.056) and ADDG09098344805.E in the respiratory tract (0.061; appendix pp 13, 27). No permissive mutation effects were predicted for rs894063168 and its linkage disequilibrium variants.

The rs13376485 allele of *FCGR2B* was not significantly associated in the subtype analyses (figure 1). However, the frequency of the risk alternative (A) allele was similar across the IgG4-related disease (0.147), Mikulicz's disease (0.132), and autoimmune pancreatitis cohorts (0.158) and lower than in controls (0.232), with consistent effect sizes of OR 0.59 [95% CI 0.51–0.69] for IgG4-related disease, 0.54 [0.42–0.68] for Mikulicz's disease, and 0.66 [95% CI 0.54–0.80] for autoimmune pancreatitis. This finding probably reflects limited statistical power due to smaller sample sizes of patients with Mikulicz's disease and autoimmune pancreatitis (appendix p 28).

In analysing the association of *HLA* with IgG4-related disease, we prioritised amino acid residues over individual single nucleotide variants. *HLA* alleles were determined for all individuals, and their amino acid sequences were aligned per gene. All alleles were retained regardless of frequency, because even rare alleles can contribute to amino acid residue-level

variation relevant to association analysis (appendix p 14). The typed *HLA* alleles were subsequently mapped to their respective amino acid residues. An omnibus test of 228 amino acid positions identified the strongest association at position seven of the β domain in the peptide-binding groove of HLA-DRB1 (referred to as DRB1-GB-7), with $p=4.9\times 10^{-37}$. The stepwise conditional analysis identified two further significant residues: A-GA2-9 ($p=1.2\times 10^{-8}$) and DQB1-GB-82 ($p=1.1\times 10^{-4}$; appendix p 15).

Multiple logistic regression analysis revealed that valine at DRB1-GB-7 was associated with an increased risk for IgG4-related disease (OR 1.86 [95% CI 1.62–2.12] $p=1.1\times 10^{-19}$), whereas aspartic acid at this position was protective (0.53 [0.43–0.66], $p=1.3\times 10^{-8}$). At A-GA2-9, both phenylalanine (0.76 [0.68–0.86], $p=4.1\times 10^{-6}$) and cysteine (0.40 [0.27–0.60], $p=7.9\times 10^{-6}$) were associated with a reduced risk, and alanine at DQB1-GB-82 also showed protective effects (0.70 [0.62–0.79], $p=4.7\times 10^{-9}$; table 3). In the subtype analysis, DRB1-GB-7 and A-GA2-9 remained as associated candidates in the autoimmune pancreatitis cohort following stepwise conditioning, whereas DRB1-GB-7 and A-GA2-62 were identified in the Mikulicz's disease cohort (appendix pp 16, 29). Arginine at A-GA2-62 was associated with an increased risk (1.39 [1.15–1.66], $p=4.7\times 10^{-4}$) compared with the reference histidine. No significant heterogeneity was detected between Set 1 and Set 2 for these HLA amino acid residues (appendix p 30).

We estimated total complement *C4* copy number and copy numbers of its isotypes (*C4A* and *C4B*) from WGS data in 869 cases of IgG4-related disease and 2659 controls. The calibrated *C4* copy number exhibited a multimodal distribution, with peaks at the integer values 2–8 (appendix p 17). By differentiating *C4A* and *C4B* through their ratios across five isoform-specific nucleotide variants, the raw *C4A* fractions aligned with the expected ratio for specific *C4* copy numbers (appendix pp 17–18). Additionally, *C4A* and *C4B* copy numbers exhibited negative correlation ($r=-0.49$).

The χ^2 goodness-of-fit test revealed significant differences in the distributions of both *C4A* and *C4B* copy numbers between IgG4-related cases and controls ($p=9.4\times 10^{-7}$ for *C4A*; $p=3.2\times 10^{-8}$ for *C4B*; figure 2A, B). In contrast, statistical significance was attenuated when analysing total *C4* copy number, defined as the sum of *C4A* and *C4B* ($p=0.0051$; figure 2C). In logistic regression models evaluating the association between *C4* copy number variations and IgG4-related disease, total *C4* copy number alone was not significantly associated with disease risk. However, a combined model including both *C4A* and *C4B* provided the best fit, as indicated by the lowest delta Akaike Information Criterion (appendix p 31). In this model, *C4A* was associated with decreased risk ($\beta=-0.127$, $SE=0.048$, $p=7.9\times 10^{-3}$), and *C4B* showed a positive association ($\beta=0.151$, $SE=0.064$, $p=1.9\times 10^{-2}$). The estimated ORs based on *C4A/C4B* copy number combinations ranged from 0.52 (95% CI 0.38–0.71) in individuals with six *C4A* and one *C4B* copies to 1.54 (1.18–1.99) in individuals with one *C4A* and four *C4B* copies (figure 2D; appendix p 32)

The *C4* gene, positioned between *HLA class I* and *HLA class II* regions on chromosome 6, was assessed for correlation between the copy number of *C4* and its isotypes with the three HLA amino acid residues significantly associated with IgG4-related disease. We obtained the highest correlation between *C4A* copy number variations and DQB1-82-Ala with Pearson's $r^2=0.15$ (appendix p 19). Both *C4A* copy number and DQB1-82-Ala remained significantly associated with IgG4-related disease after adjusting for their genetic correlation (appendix p 33).

Discussion

Building on our earlier identification of *HLA-DRB1* and *FCGR2B* as genetic determinants of IgG4-related disease,⁷ we used advanced sequencing techniques to validate and expand upon these associations. Our key findings included novel contributions from *C4A* and *C4B* copy numbers and independent roles for *HLA-A* and *HLA-DQB1*, highlighting the complex genetic architecture of IgG4-related disease. Subtype-specific analyses further uncovered additional loci, emphasising the heterogeneity of genetic risk factors across disease subtypes. A previous

GWAS in a Japanese cohort identified *IL1R1* as being specifically associated with the periaortitis and periarthritis subtype of IgG4-related disease.²⁸ Our findings reinforce the need for functional studies to elucidate mechanisms underlying disease susceptibility.

For the newly detected HLA amino acid residues, the corresponding *HLA* allele frequencies showed strong correlations with those from the previous dataset typed via SNP2HLA (Pearson's $r=0.996$ in 597 cases; Pearson's $r=0.981$ in 295 controls), and high concordance of *HLA* alleles was observed in overlapping samples (97.4%, $n=892$), indicating no substantial methodological discrepancies. Including the top ten principal components as covariates and observing no inflation in the GWAS results for both studies supports the avoidance of false positives in the *HLA* variant analysis. In our previous GWAS,⁷ controls were sampled from a narrow region (Nagahama City, Japan), resulting in slight separation of principal components between cases and controls. In contrast, broader sampling of the Japanese population in the present study improved the alignment of principal component distributions, potentially mitigating confounding effects and explaining previously diminished associations. At the *HLA* allele level, some variants corresponding to the newly detected amino acid residues became significant and even changed their direction of association. Therefore, these novel associations are likely attributable to the increased control size, more balanced population sampling, and improved accuracy in typing rare variants in the whole-genome sequencing dataset.

The DRB1-GB-7, A-GA2-9, and DQB1-GB-82 residues are located in the peptide-binding grooves of their respective HLA molecules, with DRB1-GB-7 and A-GA2-9 positioned at the centre and DQB1-GB-82 at the terminal of the α -helix (appendix p 15). Therefore, residues at these positions might play a role in influencing antigenic peptide interactions with CD4⁺ and CD8⁺ T cells via the T-cell receptor, potentially modulating autoimmune susceptibility by altering the peptide spectrum.^{29,30} Our findings suggest that HLA-A, DRB1, and DQB1 might be involved in the aetiology of IgG4-related disease, potentially through their effects on antigen

presentation efficiency and protein stability.^{31,32} These findings reinforce the need for further structural and functional studies to elucidate their contributions to disease susceptibility.

The *PTCH1* gene, which encodes the patched-1 receptor involved in repressing the Sonic hedgehog signalling pathway, has been implicated in embryonic development and tumourigenesis.³³ *PTCH1* dysregulation has been implicated in several malignancies, including colorectal, reproductive, and basal cell cancers, emphasising its importance in tumourigenesis.^{34,35} Both MENTR and eQTL analyses showed *PTCH1* expression changes attributable to rs66917126, suggesting a possible role for *PTCH1* in immune regulation. The functions and target genes of the two enhancers upstream of *PTCH1* remain unknown, and further analyses are needed to elucidate their relevance to Mikulicz's disease.

The complement system is critical for clearing cellular debris and regulating inflammation. A negative correlation between *C4A* and *C4B* was reported in a previous study of a European population.³⁷ We showed that a lower *C4A* copy number and a higher *C4B* copy number were associated with an increased risk of IgG4-related disease. The protective effect of *C4A* aligned with findings in studies on systemic lupus erythematosus and primary Sjögren's syndrome, whereas *C4B* demonstrated an opposing impact.^{37,38} Other opposing genetic associations have also been observed in IgG4-related disease and systemic lupus erythematosus. Although *Fcgr2b*-deficient mice are prone to developing systemic lupus erythematosus, increased cell-surface FCGR2B levels elevate IgG4-related disease risk.^{39,40} Similarly, *DRB1-GB-7-Val*, associated with IgG4-related disease, has been reported to be protective against systemic lupus erythematosus.⁴¹ Notably, these two diseases share the same gene determinants but exhibit opposing effects. Integration of HLA-HD results with *C4* copy number variations predictions revealed low linkage disequilibrium ($r^2 < 0.15$) among *C4A*, *C4B*, and HLA amino acid residues in the Japanese population. This observation suggests that structural changes in the peptide-binding groove from HLA amino acid residues, alongside proportional differences in *C4* isotypes, might collectively contribute to the development of IgG4-related disease.

To date, the diagnosis and classification of IgG4-related disease have predominantly relied on elevated serum IgG4 concentrations and histopathological assessments in the affected organs. However, our subtype analysis revealed that IgG4-related disease manifestations across different organs might involve distinct single nucleotide variants and HLA amino acid residues. Differences in the associations of rs66917126 and HLA-A-GA2-62 across IgG4-related disease subtypes might reflect the distinct genetic pathways in the pathogenesis of Mikulicz's disease compared with IgG4-related disease and autoimmune pancreatitis. Further studies exploring the molecular mechanisms underlying these associations could improve our understanding of immune dysregulation across subtypes and refine the classification of IgG4-related disease.

This study is limited by its small sample size, reflecting the rarity of IgG4-related disease and recruitment challenges, which also precluded adequately powered sex-stratified analyses. Despite the rising prevalence in Japan due to improved awareness, the true prevalence might still be underestimated due to diagnostic difficulties. In the subtype-specific analyses, patients were not strictly classified into mutually exclusive subtypes. IgG4-related disease often presents with overlapping subtypes, and full systemic evaluations are not always conducted in clinical settings. As a result, some coexisting subtypes might not have been documented. To address this limitation, the development of a more systematic diagnostic framework will be essential in future studies. In whole-genome sequencing analyses using short-read sequencing, achieving high mapping rates in regions such as multigene families remains challenging due to their highly repetitive nature, which limits analytical resolution. The *HLA* and *FCGR* loci are representative of such regions. *HLA* coding sequences were resolved using unmapped short reads and the HLA-HD algorithm, allowing for association analyses of alleles and amino acid residues. However, the non-coding regions of *HLA* were excluded due to low mapping quality, despite their potential relevance to disease susceptibility. Furthermore, single nucleotide variants surrounding *FCGR2B*, including rs1340976, which is associated with IgG4-related disease,⁷ and rs1050501, which is associated systemic lupus erythematosus,

were also excluded from the present analysis. Moreover, short-read sequencing is insufficient to resolve haplotypes of copy number variation and human endogenous retroviral insertions into specific *C4* isotypes.⁴² A comprehensive analysis of such complexed gene loci using long-read sequencing approaches will be necessary to address this question. Additionally, interrelating expression quantitative trait loci, protein, and metabolite data is essential to deepen our understanding of the pathogenesis of IgG4-related disease. The blood concentrations of C4 and its two isoforms have not been examined, and their involvement in IgG4-related disease remains to be elucidated. Future proteomic analyses using patient blood samples might help clarify their role in the disease pathology. We also anticipate studies in diverse ancestral populations to compare the shared or distinct genomic features of IgG4-related disease.

The Japanese IgG4-Related Disease Working Consortium

Kyoto Medical Center, Kyoto, Japan (Akira Shimatsu); Osaka Habikino Medical Center, Osaka, Japan (Atsushi Ogata); Kagoshima University, Kagoshima, Japan (Hirohito Tsubouchi); Saiseikai Ibaraki Hospital, Osaka, Japan (Hiroshi Tatsuta); Tokai University Hachioji Hospital, Tokyo, Japan (Hitoshi Ichikawa); Shimane University Hospital, Shimane, Japan (Ichiro Moriyama); Nagoya City University Midori Municipal Hospital, Nagoya, Japan (Itaru Naitoh); Kobe City Nishi-Kobe Medical Center, Kobe, Japan (Jun Mimura); Fujita Health University Bantane Hospital, Nagoya, Japan (Kazuo Inui); Keio University, Tokyo, Japan (Kazuo Tsubota and Takanori Kanai); Shiga General Hospital, Shiga, Japan (Kazuyoshi Matsumura); Kyushu Medical Center, Fukuoka, Japan (Ken Kawabe); Kishiwada City Hospital, Osaka, Japan (Kozo Kajimura); Tenri Hospital, Nara, Japan (Masaya Ohana); Takamatsu Red Cross Hospital, Kagawa, Japan (Mitsushige Shibatoge); Takatsuki Red Cross Hospital, Osaka, Japan (Naoki Kanda); Nagano Municipal Hospital, Nagano, Japan (Osamu Hasebe); Nagoya University, Nagoya, Japan (Shigeo Nakamura); Tokyo Women's Medical University Yachiyo Medical Center, Chiba, Japan (Takayoshi Nishino); Kansai Electric Power Hospital, Osaka, Japan (Takefumi Nakamura); Kobe City Medical Center General Hospital, Kobe, Japan (Tetsuro Inokuma); Kyoto-Katsura Hospital, Kyoto, Japan (Toshihiro Kusaka); and Japanese Red Cross Wakayama Medical Center, Wakayama, Japan (Yukitaka Yamashita).

Contributors

FM, MKaw, SN, TM, KOk, TC, and HU contributed to the funding acquisition. FM and TC designed and supervised the study. IY, TKaw, MN, and MT contributed to data curation. YOZ and TIw developed the methodology. YOZ and TIw conducted informatic analysis. YOZ, TIw, SK, and FM wrote the manuscript, and all authors critically revised and approved the final version. HT, AK, KK, HD, MY, MS, TIk, SM, MKan, KN, KY, AA, YM, IM, YKu, HS, TO, SY, KH, AM, NM, HSh, MM, YH, YKo, TS, TKin, CK, MA, TA, AN, KOs, YT, HY, TKam, TKim, HI, and the Japanese IgG4-Related Disease Working Consortium made clinical diagnosis of the patients and collected their clinical information and biological materials. FM, SK, TKaw, IY, and YOZ

directly assessed and verified the underlying data. All authors had full access to the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

NM reports grants or contracts from Novartis, MSD, Incyte, Ono Pharmaceutical, Seagen, Boehringer Ingelheim, Pfizer, AstraZeneca, and SERVIER (to the institution); honoraria from Yakult Honsha, AstraZeneca, Novartis, FUJIFILM Toyama Chemical, Taiho Pharmaceutical, Miyarsan Pharmaceutical, SERVIER, and Teijin Pharma; and participation on advisory boards for MSD, Boehringer Ingelheim, and Novartis. YM reports scholarship fund donations from Chugai Pharma, Kyowa Hakko Kirin Pharma, and Otsuka Pharma. YT reports personal fees (speaking fees and honoraria) from Chugai Pharma, UCB, AbbVie, AstraZeneca, Eli-Lilly, Behringer-Ingelheim, GSK, Eisai, IQVIA, Daiichi-Sankyo, Otsuka Pharma, Taisho, Gilead, and Bristol Myers Squibb. All other authors declare no conflicts of interest.

Data sharing

The full GWAS summary statistics are available at <https://www.hgvd.genome.med.kyoto-u.ac.jp/repository/HGV0000023.html>.

Acknowledgments

This study was supported in part by the Japanese Ministry of Health, Labour, and Welfare (grant number H24-Nanchi-011, 23FC1015), the Japanese Agency of Medical Research and Development (grant numbers JP17EK0109283, JP16EK0109196, JP18EK0109348, JP21EK0109548, JP23EK0109675, and JP24EK0109759), the Kyoto University Grant for Top Global University Japan Project, and the Kyoto University Division of Graduate Studies SPRING Program. We thank Miki Kokubo for sample management and technical assistance for the genomic study. We also express our gratitude to all the patients for their contributions.

Reference

- 1 Umehara H, Okazaki K, Masaki Y, et al, and the Research Program for Intractable Disease by Ministry of Health, Labor and Welfare (MHLW) Japan G4 team. A novel clinical entity, IgG4-related disease (IgG4RD): general concept and details. *Mod Rheumatol* 2012; **22**: 1–14.
- 2 Hamano H, Kawa S, Horiuchi A, et al. High serum IgG4 concentrations in patients with sclerosing pancreatitis. *N Engl J Med* 2001; **344**: 732–38.
- 3 Peyronel F, Della-Torre E, Maritati F, et al. IgG4-related disease and other fibro-inflammatory conditions. *Nat Rev Rheumatol* 2025; **21**: 275–90.
- 4 Uchida K, Masamune A, Shimosegawa T, Okazaki K. Prevalence of IgG4-related disease in Japan based on nationwide survey in 2009. *Int J Rheumatol* 2012; **2012**: 358371.
- 5 Masamune A, Kikuta K, Hamada S, et al, and the Collaborators. Nationwide epidemiological survey of autoimmune pancreatitis in Japan in 2016. *J Gastroenterol* 2020; **55**: 462–70.
- 6 Wallace ZS, Miles G, Smolkina E, et al. Incidence, prevalence and mortality of IgG4-related disease in the USA: a claims-based analysis of commercially insured adults. *Ann Rheum Dis* 2023; **82**: 957–62.
- 7 Terao C, Ota M, Iwasaki T, et al, and the Japanese IgG4-Related Disease Working Consortium. IgG4-related disease in the Japanese population: a genome-wide association study. *Lancet Rheumatol* 2019; **1**: e14–22.
- 8 Yang Y, Chung EK, Wu YL, et al. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 2007; **80**: 1037–54.

- 9 Umehara H, Okazaki K, Kawa S, et al, and the Research Program for Intractable Disease by the Ministry of Health, Labor and Welfare (MHLW) Japan. The 2020 revised comprehensive diagnostic (RCD) criteria for IgG4-RD. *Mod Rheumatol* 2021; **31**: 529–33.
- 10 Higasa K, Kamatani Y, Kawaguchi T, et al. Whole-genome sequencing of 3,135 individuals representing the genetic diversity of the Japanese population. medRxiv 2025; 10.02.25336626. <https://www.medrxiv.org/content/10.1101/2025.10.02.25336626v1>.
- 11 Kawai Y, Hitomi Y, Ueta M, et al. Mapping of susceptible variants for cold medicine-related Stevens-Johnson syndrome by whole-genome resequencing. *NPJ Genom Med* 2021; **6**: 9.
- 12 Shimosegawa T, Chari ST, Frulloni L, et al, and the International Association of Pancreatology. International consensus diagnostic criteria for autoimmune pancreatitis: guidelines of the International Association of Pancreatology. *Pancreas* 2011; **40**: 352–58.
- 13 Nagasaki M, Sekiya Y, Asakura A, et al. Design and implementation of a hybrid cloud system for large-scale human genomic research. *Hum Genome Var* 2023; **10**: 6.
- 14 Sakaue S, Hirata J, Kanai M, et al. Dimensionality reduction reveals fine-scale structure in the Japanese population with consequences for polygenic risk prediction. *Nat Commun* 2020; **11**: 1569.
- 15 Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.
- 16 Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* 2016; **32**: 1423–26.
- 17 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**: 2190–91.

- 18 Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 2016; **32**: 1493–501.
- 19 Koido M, Hon C-C, Koyama S, et al. Prediction of the cell-type-specific transcription of non-coding RNAs from genome sequences via machine learning. *Nat Biomed Eng* 2023; **7**: 830–44.
- 20 Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. HLA-HD: an accurate *HLA* typing algorithm for next-generation sequencing data. *Hum Mutat* 2017; **38**: 788–97.
- 21 Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012; **44**: 291–96.
- 22 Lefranc M-P, Duprat E, Kaas Q, Tranne M, Thiriot A, Lefranc G. IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev Comp Immunol* 2005; **29**: 917–38.
- 23 Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004; **25**: 1605–12.
- 24 Belt KT, Carroll MC, Porter RR. The structural basis of the multiple forms of human complement component C4. *Cell* 1984; **36**: 907–14.
- 25 McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–303.
- 26 Babadi M, Fu JM, Lee SK, et al. GATK-gCNV enables the discovery of rare copy number variants from exome sequencing data. *Nat Genet* 2023; **55**: 1589–97.

- 27 Vösa U, Claringbould A, Westra H-J, et al, and the BIOS Consortium, and the i2QTL Consortium. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* 2021; **53**: 1300–10.
- 28 Ashihara N, Ota M, Fujinaga Y, et al. The levels of IL-1 β and soluble IL-1 receptors in patients with IgG4-related periaortitis/periarteritis. *Adv Med Sci* 2022; **67**: 257–61.
- 29 Scally SW, Petersen J, Law SC, et al. A molecular basis for the association of the *HLA-DRB1* locus, citrullination, and rheumatoid arthritis. *J Exp Med* 2013; **210**: 2569–82.
- 30 van Lummel M, Duinkerken G, van Veelen PA, et al. Posttranslational modification of HLA-DQ binding islet autoantigens in type 1 diabetes. *Diabetes* 2014; **63**: 237–47.
- 31 Jin H, Arase N, Hirayasu K, et al. Autoantibodies to IgG/HLA class II complexes are associated with rheumatoid arthritis susceptibility. *Proc Natl Acad Sci USA* 2014; **111**: 3787–92.
- 32 Miyadera H, Ohashi J, Lernmark Å, Kitamura T, Tokunaga K. Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA. *J Clin Invest* 2015; **125**: 275–91.
- 33 Ming JE, Kaupas ME, Roessler E, et al. Mutations in *PATCHED-1*, the receptor for SONIC HEDGEHOG, are associated with holoprosencephaly. *Hum Genet* 2002; **110**: 297–301.
- 34 Wang Y, Chen H, Jiao X, et al. *PTCH1* mutation promotes antitumor immunity and the response to immune checkpoint inhibitors in colorectal cancer patients. *Cancer Immunol Immunother* 2022; **71**: 111–20.
- 35 Yang X-H, Xu B-H, Zhou D-L, et al. Inherited rare and common variants in *PTCH1* and *PTCH2* contributing to the predisposition to reproductive cancers. *Gene* 2022; **814**: 146157.
- 36 Chen L, Ge B, Casale FP, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* 2016; **167**: 1398–1414.e24.

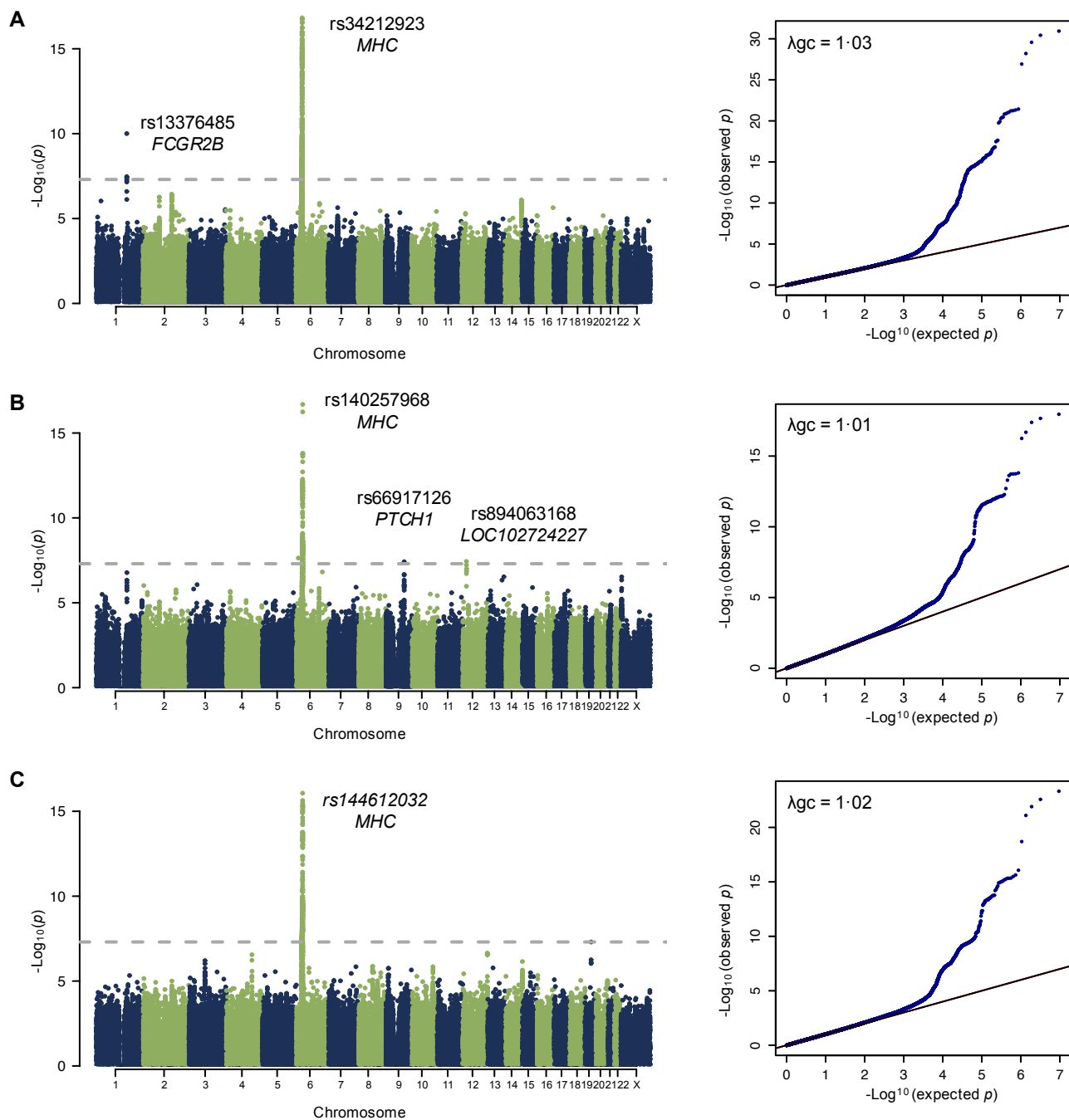
- 37 Lundtoft C, Pucholt P, Martin M, et al, and the DISSECT Consortium, and the ImmunoArray Development Consortium. Complement *C4* copy number variation is linked to SSA/Ro and SSB/La autoantibodies in systemic inflammatory autoimmune diseases. *Arthritis Rheumatol* 2022; **74**: 1440–50.
- 38 Kamitaki N, Sekar A, Handsaker RE, et al, and the Schizophrenia Working Group of the Psychiatric Genomics Consortium. Complement genes contribute sex-biased vulnerability in diverse disorders. *Nature* 2020; **582**: 577–81.
- 39 Kyogoku C, Dijstelbloem HM, Tsuchiya N, et al. Fcγ receptor gene polymorphisms in Japanese patients with systemic lupus erythematosus: contribution of *FCGR2B* to genetic susceptibility. *Arthritis Rheum* 2002; **46**: 1242–54.
- 40 Zhao M, Wigren M, Dunér P, et al. FcγRIIB inhibits the development of atherosclerosis in low-density lipoprotein receptor-deficient mice. *J Immunol* 2010; **184**: 2253–60.
- 41 Kim K, Bang S-Y, Lee H-S, et al. The HLA-DRβ1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. *Nat Commun* 2014; **5**: 5902.
- 42 Dangel AW, Mendoza AR, Baker BJ, et al. The dichotomous size variation of human complement *C4* genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* 1994; **40**: 425–36.

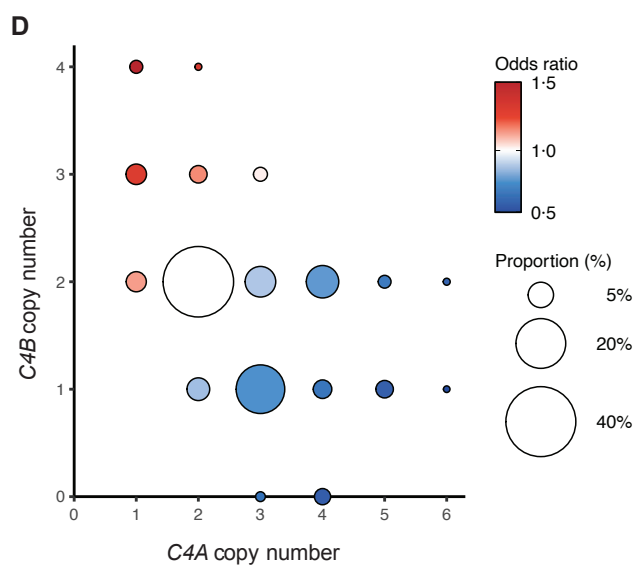
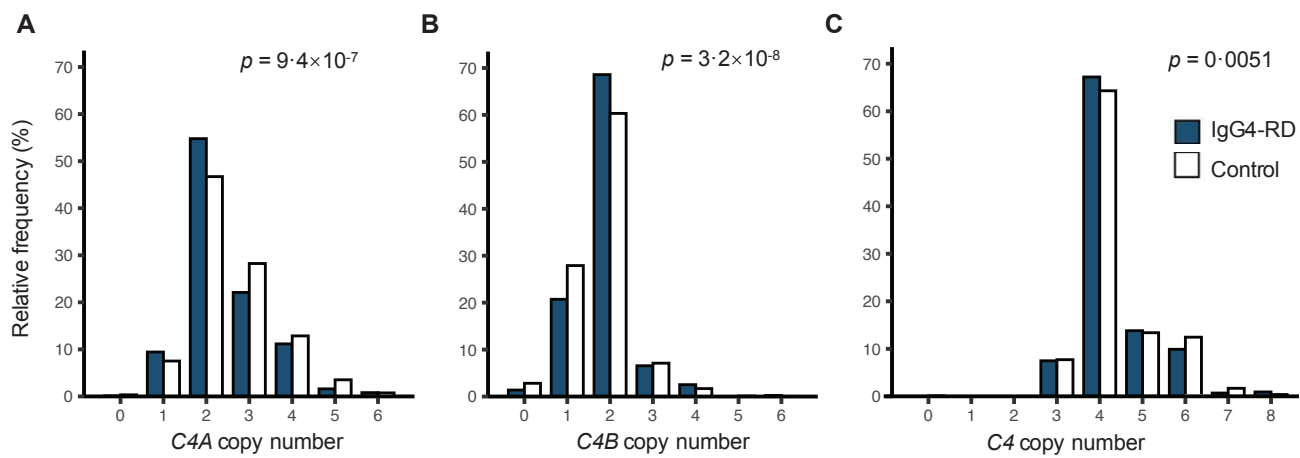
Figure 1: Genome-wide association analysis of IgG4-related disease and its subtypes

Manhattan plots (left) and quantile–quantile plots (right) for all cases with IgG4-related disease (A), Mikulicz’s disease (B), and autoimmune pancreatitis (C). In the Manhattan plots, the x-axis represents chromosomal positions, and the y-axis shows the $-\log_{10}(\text{p value})$ from association tests. The horizontal grey dashed line indicates the genome-wide significance threshold ($p=5.0 \times 10^{-8}$). Significantly associated single nucleotide variants are labelled with their reference single nucleotide variant identification numbers and corresponding gene names. In the quantile–quantile plots, the x-axis represents the expected $-\log_{10}(\text{p value})$ under the null hypothesis, and the y-axis shows the observed $-\log_{10}(\text{p value})$. Each dot represents an individual single nucleotide variant, and the diagonal line corresponds to the null distribution. The λ_{gc} is indicated for each analysis. λ_{gc} =genomic inflation factor.

Figure 2: Association between C4 copy number variation and IgG4-related disease

Distributions of *C4A* (A), *C4B* (B), and total *C4* (C) copy numbers in cases with IgG4-related disease and controls. The x-axis indicates the copy number of each gene, and the y-axis represents the relative frequency (%) in cases and controls. (D) Combined effects of *C4A* and *C4B* copy numbers on the risk of IgG4-related disease. Each circle represents individuals with a specific combination of *C4A* and *C4B* copy numbers. Circle size indicates the proportion of control individuals in the cohort with that genotype. Circle colour reflects the odds ratio, according to the accompanying scale bar, with red indicating increased risk and blue indicating decreased risk.





	Set 1		Set 2	
	Case (n=646)*	Control (n=2254)*	Case (n=223)*	Control (n=405)*
Age at diagnosis, years	64·4 (11·4)	..	63·5 (10·9)†	..
Sex				
Female	172 (26·6%)	1348 (59·8%)	78 (35·0%)	65 (16·0%)
Male	474 (73·4%)	906 (40·2%)	145 (65·0%)	340 (84·0%)
Ethnicity				
Japanese of Hondo ancestry	646 (100%)	2254 (100%)	223 (100%)	405 (100%)
IgG4 concentration at diagnosis, mg/dL	653·1 (596·3)	..	543·5 (603·5)‡	..
Diagnosis of IgG4-related disease§				
Definite	603 (93·3%)	..	165 (74·0%)	..
Probable	25 (3·9%)	..	6 (2·7%)	..
Possible	18 (2·8%)	..	52 (23·3%)	..
Disease subtype				
Autoimmune pancreatitis only	322 (49·8%)	..	33 (14·8%)	..
Mikulicz's disease only	196 (30·3%)	..	65 (29·1%)	..
Autoimmune pancreatitis and Mikulicz's disease	72 (11·1%)	..	30 (13·5%)	..
Other subtypes or subtype unknown	56 (8·7%)	..	95 (42·6%)	..

Data are mean (SD), or n (%). *The number included in the association studies after the standard quality control procedure. †Calculated based on 192 patients with a confirmed age at diagnosis. ‡Calculated based on 221 patients with confirmed IgG4 concentrations at diagnosis. §According to the revised comprehensive diagnostic criteria for IgG4-related disease in 2020.¹¹

Table 1: Demographic and clinical characteristics of the study population

Chromosomal positions and reference and alternative alleles of the single-nucleotide variants are according to the GRCh38 reference genome. OR=odds ratio.

	Chromosome: position	Gene	Allele, reference/alternative	Alternative allele frequency			OR (95% CI)	p value
				Case	Control	All		
IgG4-related disease								
rs13376485	1: 161676087	FCGR2B	G/A	0.147	0.232	0.211	0.59 (0.51–0.69)	9.8×10 ^{−11}
rs7539053	1: 161690055	FCGR2B	T/C	0.525	0.610	0.589	0.71 (0.63–0.80)	3.4×10 ^{−8}
rs56374312	1: 161695965	FCGR2B	C/T	0.095	0.156	0.141	0.58 (0.48–0.70)	3.6×10 ^{−8}
rs6668534	1: 161702205	FCRLA	G/A	0.527	0.611	0.590	0.72 (0.64–0.81)	4.4×10 ^{−8}
Mikulicz’s disease								
rs66917126	9: 95560433	PTCH1	C/T	0.159	0.094	0.102	1.94 (1.53–2.46)	3.8×10 ^{−8}
rs894063168	12: 18817706	LOC102724227	C/T	0.018	0.002	0.004	12.57 (5.11–30.93)	3.6×10 ^{−8}

Table 2: Genetic variants showing genome-wide significance for IgG4-related disease and subtypes

	Amino acid frequency			p value	OR (95% CI)	Corresponding <i>HLA</i> alleles*
	Case	Control	All			
DRB1-GB-7						
Ser	0.334	0.377	0.366	Reference	..	*08:02, *08:03, *11:01, *12:01, *12:02, *13:02, *14:03, *14:05, *14:06, *14:54
Val	0.378	0.228	0.265	1.1×10 ⁻¹⁹	1.86 (1.62–2.12)	*04:03, *04:05, *04:06, *04:10
Asp	0.064	0.139	0.121	1.3×10 ⁻⁸	0.53 (0.43–0.66)	*09:01
Leu	0.055	0.067	0.064	0.58	0.93 (0.73–1.19)	*01:01
Pro	0.169	0.189	0.184	0.86	1.01 (0.86–1.20)	*15:01, *15:02
A-GA2-9						
Tyr	0.656	0.581	0.600	Reference	..	*02:01, *02:06, *26:01, *26:02, *26:03, *31:01, *33:03
Phe	0.328	0.382	0.369	4.1×10 ⁻⁶	0.76 (0.68–0.86)	*24:02
Cys	0.016	0.036	0.031	7.9×10 ⁻⁶	0.40 (0.27–0.60)	*02:07, *11:01
DQB1-GB-82						
Glu	0.586	0.520	0.536	Reference	..	*03:01, *03:02, *03:03, *04:01, *04:02
Ala	0.320	0.401	0.381	4.7×10 ⁻⁹	0.70 (0.62–0.79)	*05:01, *05:02, *05:03, *06:01, *06:02
Gly	0.094	0.079	0.079	0.51	1.07 (0.88–1.29)	*06:04

Bonferroni correction $p=1.53 \times 10^{-4}$. **HLA* alleles with a frequency of greater than 0.01 in either the case or control population are shown.

Table 3: Susceptible *HLA* amino acid residues and corresponding *HLA* alleles

THE LANCET

Rheumatology

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed.
We post it as supplied by the authors.

Supplement to: Zhang YO, Iwasaki T, Kawaguchi T, et al. IgG4-related disease in the Japanese population: a whole-genome sequencing study. *Lancet Rheumatol* 2025; published online Nov 3. [https://doi.org/10.1016/S2665-9913\(25\)00195-X](https://doi.org/10.1016/S2665-9913(25)00195-X).

Supplementary Appendix

Table of contents

Supplementary Method A : Detailed description of quality control steps	4
Supplementary Method B : Detailed method for the estimation of <i>C4</i> copy number.....	4
Supplementary Figure 1. Selection of quality control thresholds for GWAS.....	6
Supplementary Figure 2. Workflow for sample and variant quality control.	7
Supplementary Figure 3. The principal component analyses of the study subjects.	8
Supplementary Figure 4. CNV estimation workflow of <i>C4</i> and its two isotypes.....	9
Supplementary Figure 5. Linkage disequilibrium of disease-associated SNVs in the <i>FCGR2B</i> locus.....	10
Supplementary Figure 6. Regional plot of chromosomal regions encompassing the associated SNVs.....	11
Supplementary Figure 7. GWAS results of IgG4-related disease and its subtypes in sample sets Set 1 and Set 2.....	12
Supplementary Figure 8. Regulatory effects of SNVs associated with MD, eQTL, and MENTR.....	13
Supplementary Figure 9. Regional plot of GWAS results in the <i>MHC</i> locus.	14
Supplementary Figure 10. Association analysis of HLA amino acid residues with IgG4-related disease	15
Supplementary Figure 11. Association analysis of HLA amino acid residues with IgG4-related disease subtypes.	16
Supplementary Figure 12. Distribution of <i>C4</i> copy number and estimated fraction of <i>C4A</i> in each group.	17
Supplementary Figure 13. Distribution of <i>C4A</i> fractions stratified by <i>C4</i> copy number.....	18
Supplementary Figure 14. Correlation between <i>C4</i> CNVs and HLA amino acid residues.....	19
Supplementary Table 1. The list of the participating institutes in the Japanese IgG4-Related Disease Working Consortium.	20
Supplementary Table 2. Summary of the whole-genome sequencing results.....	22

Supplementary Table 3. Summary of SNVs significantly associated with IgG4-related disease in Set 1, Set 2, and the meta-analysis.....	23
Supplementary Table 4. Credible SNVs by fine-mapping analysis.	24
Supplementary Table 5. Heterogeneity of newly detected SNVs in different IgG4-related disease subsets.	25
Supplementary Table 6. eQTL results for the variants in Linkage Disequilibrium with rs66917126 in the <i>PTCH</i> region.....	26
Supplementary Table 7. Predicted effect of variants associated with IgG4-related disease.	27
Supplementary Table 8. Association of rs13376485 with IgG4-related disease, MD, and AIP.	28
Supplementary Table 9. HLA amino acid residues associated with IgG4-related disease subtypes.	29
Supplementary Table 10. Heterogeneity of IgG4-related disease associated HLA amino acid residues in Set 1 and Set 2.....	30
Supplementary Table 11. Results of association analyses of IgG4-related disease risk against <i>C4</i> copy number variation.	31
Supplementary Table 12. Predicted relative onset risk of IgG4-related disease based on <i>C4</i> CNV combinations.	32
Supplementary Table 13. Multi-variate logistic regression of <i>C4A</i> and HLA-DQB1-GB-82.	33

Supplementary Method A : Detailed description of quality control steps

We conducted quality control (QC) as shown in the schematic flowchart. First, the WGS reads were aligned and jointly called with those of other WGS studies. Consequently, we conducted by GATK Best Practices to remove (1) variants with low quality (variant quality score log-odds [VQSLOD]<10 or mapping quality [MQ]<58·75), (2) heterozygous genotypes with imbalanced allele ratios (allele balance [AB]<0·25), and (3) genotypes with insufficient confidence or coverage (genotype quality [GQ]<20 or depth of coverage [DP]<7 for autosomes or X chromosomes in females, DP<3 for X chromosomes in males) (Supplementary Figure 2A-C).

We then performed the first four sequential sample-based QC to remove individuals who are (1) not inferred as Hondo ancestry by PCA using BioBank Japan reference samples (Supplementary Figure 3), (2) with low genotype call rates (<96% for Set 1 and <98% for Set 2), (3) with excess heterozygosity (>±5 standard deviations from the mean) (Supplementary Figure 2D) and (4) with cryptic relatedness (PI_HAT>0·2).

Finally, variant filtering was performed by removing those with (1) low genotype call rates (<98%), (2) low minor allele counts (allele count [AC]<10 for Set 1, AC<5 for Set 2) and (3) deviation from Hardy–Weinberg equilibrium (HWE $p<1.0\times10^{-6}$). All QC procedures were applied independently to each dataset.

Supplementary Method B : Detailed method for the estimation of *C4* copy number

A group of 80 Utah residents with Northern and Western European ancestry from the CEPH collection (CEU), whose *C4* copy numbers were determined via quantitative polymerase chain reaction (qPCR), was used for batch calibration.¹ High-coverage WGS data for this calibration group were obtained from the 1000 Genomes Project.²

The *C4* copy number was estimated using the GermlineCNVCaller (gCNV) tool from the Genome Analysis Toolkit (GATK, version 4.2.6.1), leveraging read-depth information from the WGS data.^{3,4} In gCNV's cohort mode, which improves CNV detection sensitivity and specificity by incorporating multiple samples, the recommended maximum batch size is 300 samples. To ensure consistency, study subjects were divided into batches of 300, comprising 220 test samples and a fixed set of 80 CEU samples, to standardise results across batches. Within each batch, the known *C4* copy numbers of the calibration group were used to generate a linear regression equation. This equation was applied to adjust the predicted *C4* copy numbers of the test samples, yielding calibrated values. For each sample, reads aligned within the *C4A* and *C4B* gene regions ±5 kilobases (kb) were extracted using Samtools (version 1.13)⁵ and remapped with BWA (version 0.7.13)⁶ to a reference sequence in which the *C4A* gene ±1 kb was masked. This approach ensured that reads aligning to the *C4B* gene region represented the overall *C4* gene coverage, with those carrying *C4A*-specific sequence considered as alternative alleles relative to the *C4B* reference.

The gCNV tool then determined the raw copy number for each 1,000-bp bin within the *C4* gene based on short-read depth. The mean of all bins was calculated to obtain the overall raw copy number of the *C4* gene for each individual. Among all calibrated *C4* copy numbers from the batches, k-means clustering was applied to convert the fractional *C4* copy numbers into integer values, determining the final *C4* copy numbers.⁷ The copy numbers of *C4* isotypes, *C4A* and *C4B*, were then ascertained by assessing the relative read depths covering *C4A*- and *C4B*-specific nucleotide sequences.⁸ Genetic variants in *C4A* and *C4B* were identified at base-pair resolution using the GATK HaplotypeCaller tool. The *C4A* fraction was calculated as $C4A/(C4A+C4B)$, taking the median across five SNVs. For each sample, the *C4A* fraction was rounded to the nearest plausible value within the group with a specific *C4* copy number (e.g., for a *C4* copy number of 3, possible *C4A* fractions are 0, 1/3, 2/3, or 1; a fraction of 0.29 would be rounded to 1/3). The *C4A* copy number was then calculated by multiplying the rounded *C4A* fraction by the integer *C4* copy number. Human endogenous retroviral elements inserted in the *C4* genes

were not analysed.⁹ The workflow for *C4* CNV estimation and the evaluation of batch effects are detailed (Supplementary Figure 5).

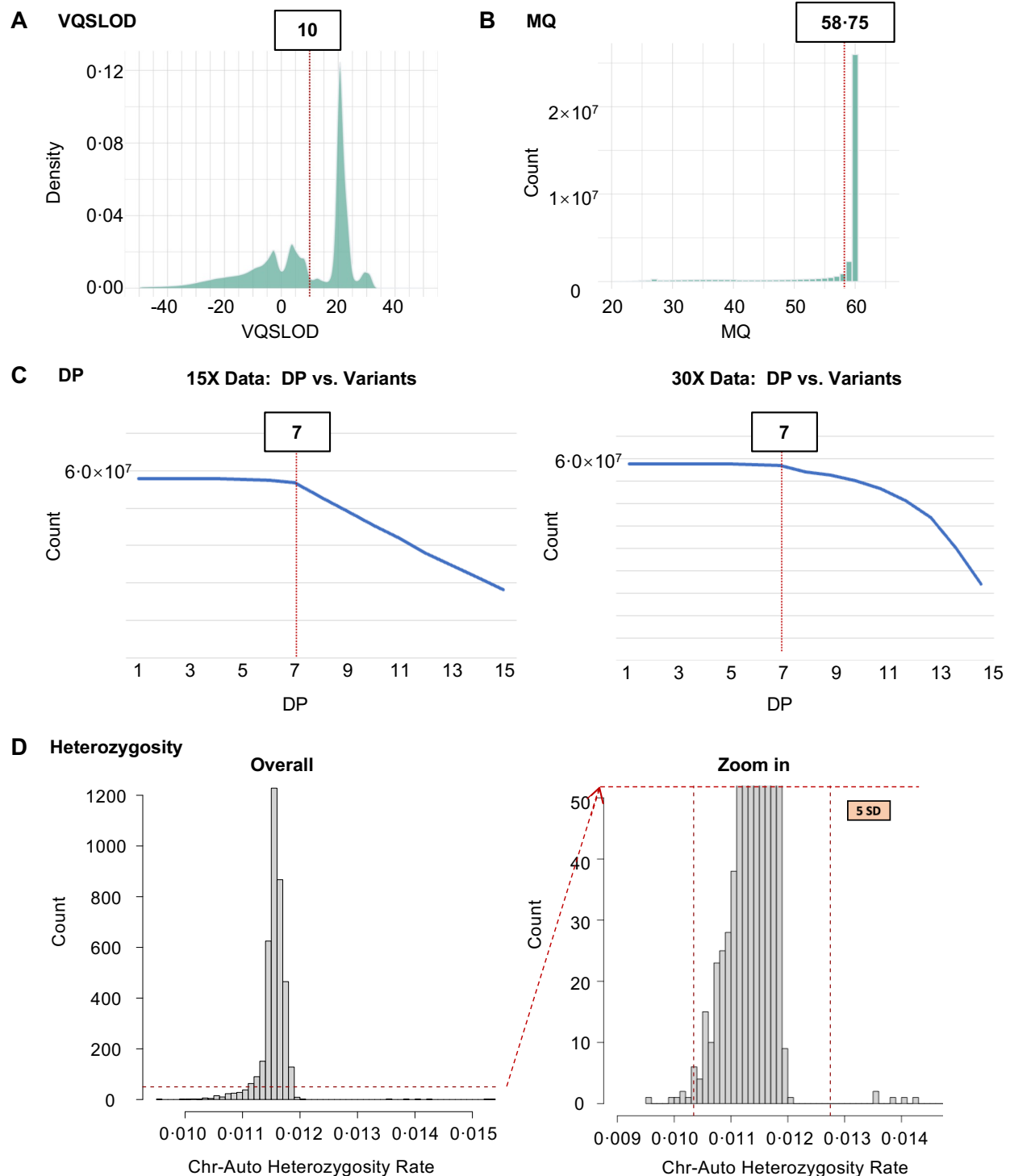
To assess differences in the distribution of *C4*, *C4A*, and *C4B* copy numbers between IgG4-related disease patients and healthy controls, the Chi-squared goodness-of-fit test was applied, with the control group considered as the theoretical distribution. To further evaluate the association between IgG4-related disease and *C4* CNV, univariate and multivariate logistic regression analyses were conducted. All statistical analyses were performed using R (version 4.1.2). Bonferroni correction was applied in both the Chi-squared goodness-of-fit tests and logistic regression models.

Supplementary Reference

- 1 Fernando MMA, Boteva L, Morris DL, *et al.* Assessment of complement C4 gene copy number using the paralog ratio test. *Hum Mutat* 2010; **31**: 866–74.
- 2 Byrska-Bishop M, Evani US, Zhao X, *et al.* High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* 2022; **185**: 3426–3440.e19.
- 3 Babadi M, Fu JM, Lee SK, *et al.* GATK-gCNV enables the discovery of rare copy number variants from exome sequencing data. *Nat Genet* 2023; **55**: 1589–97.
- 4 McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–303.
- 5 Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–9.
- 6 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–60.
- 7 MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. University of California Press, 1967: 281–98.
- 8 Belt KT, Carroll MC, Porter RR. The structural basis of the multiple forms of human complement component C4. *Cell* 1984; **36**: 907–14.
- 9 Dangel AW, Mendoza AR, Baker BJ, *et al.* The dichotomous size variation of human complement C4 genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* 1994; **40**: 425–36.

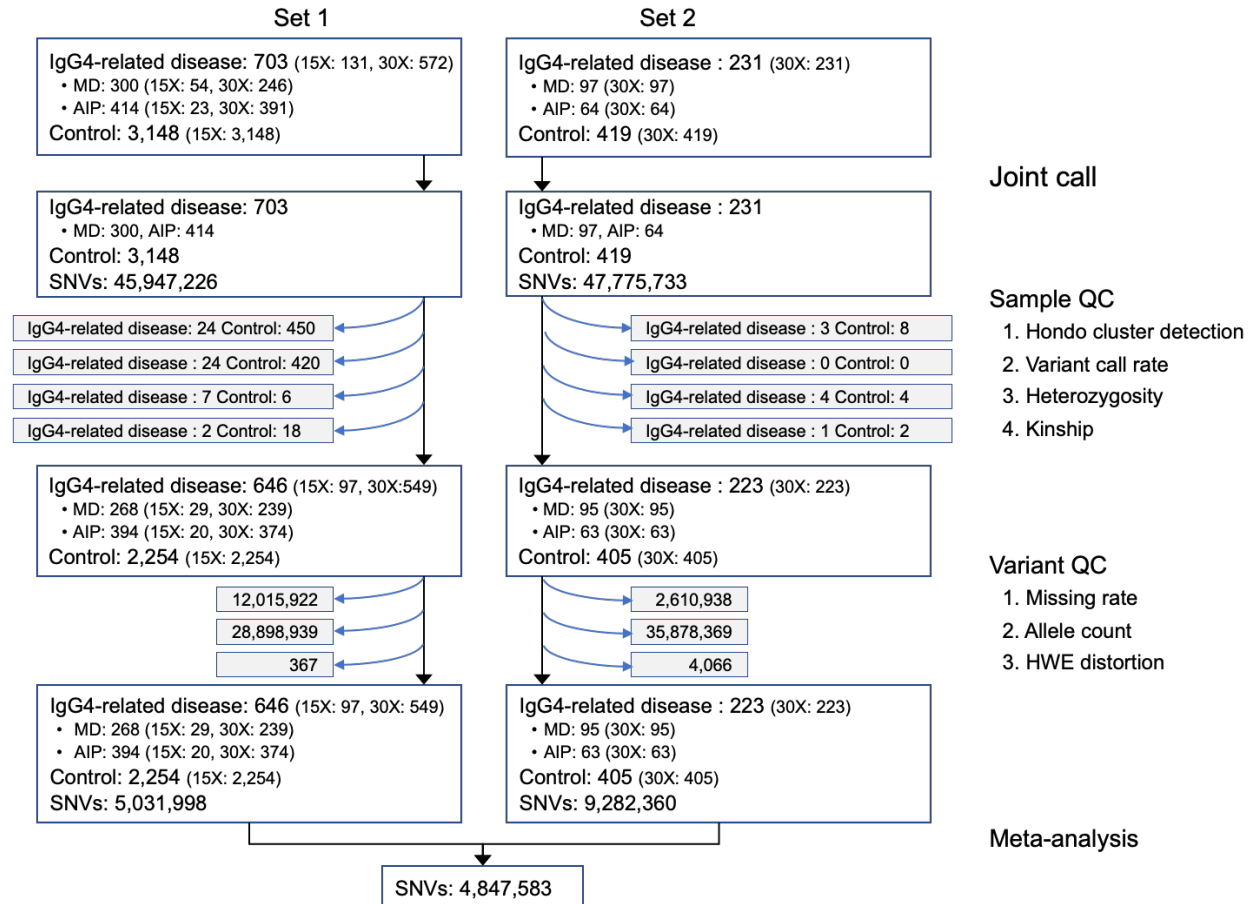
Supplementary Figure 1. Selection of quality control thresholds for GWAS.

(A) Density distribution of variant quality score log-odds (VQSLOD), with a filtering threshold of VQSLOD <10. (B) Histogram of mapping quality (MQ), with a filtering threshold of MQ <58.75. (C) Depth of coverage (DP) plotted against variant counts in 15× and 30× sequencing datasets, with a filtering threshold of DP<7. (D) Histogram of autosomal heterozygosity rates across all samples, with a zoomed-in view highlighting outliers defined as >±5 standard deviations from the mean.



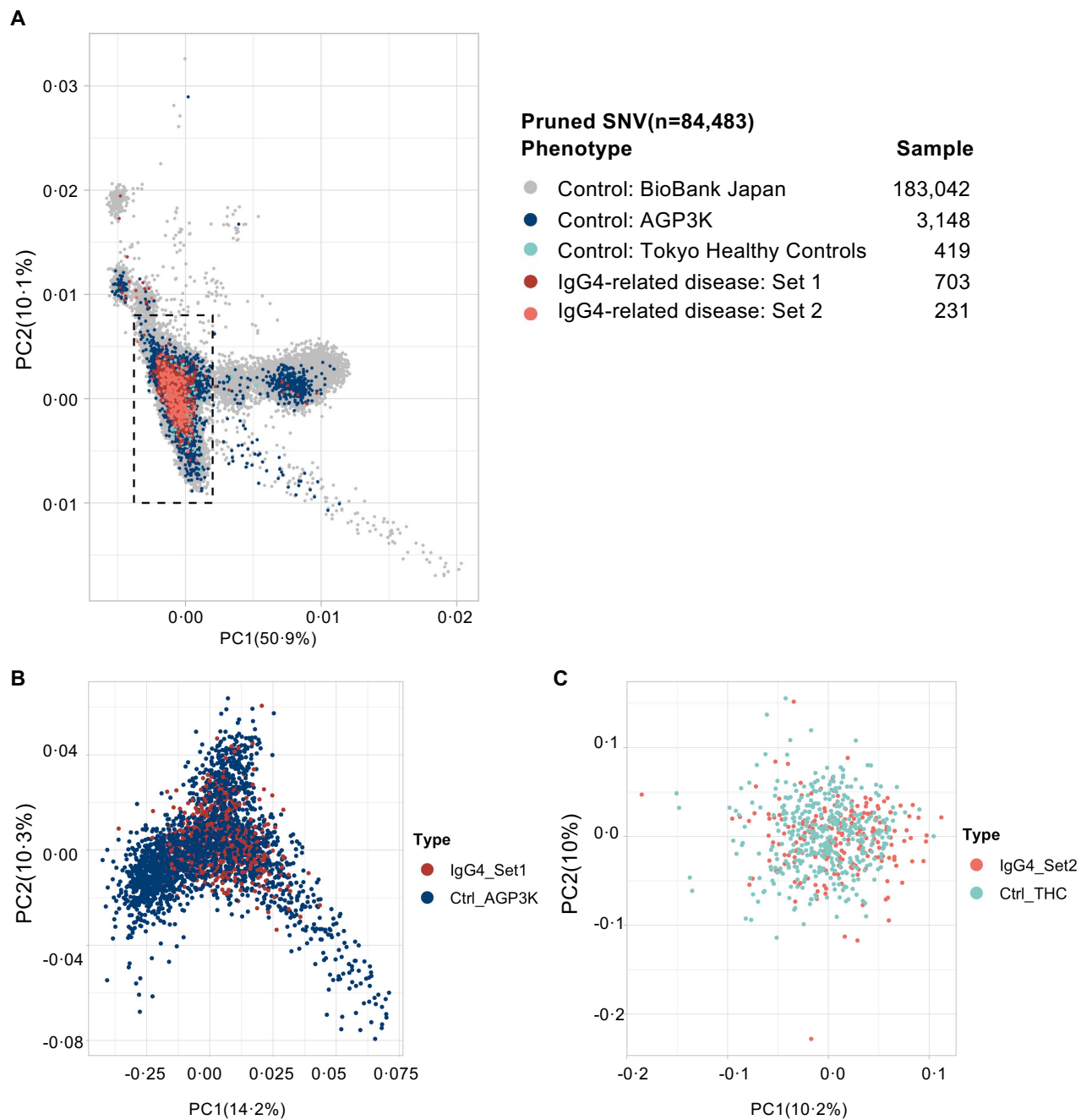
Supplementary Figure 2. Workflow for sample and variant quality control.

At each step of the procedure, indicated on the right side of the chart, the number of study subjects with coverage depth and the number of SNVs are shown in open boxes. The number of individuals and SNVs removed according to the QC criteria is shown in light grey boxes. MD: Mikulicz's disease; AIP: autoimmune pancreatitis; HWE: Hardy-Weinberg Equilibrium.



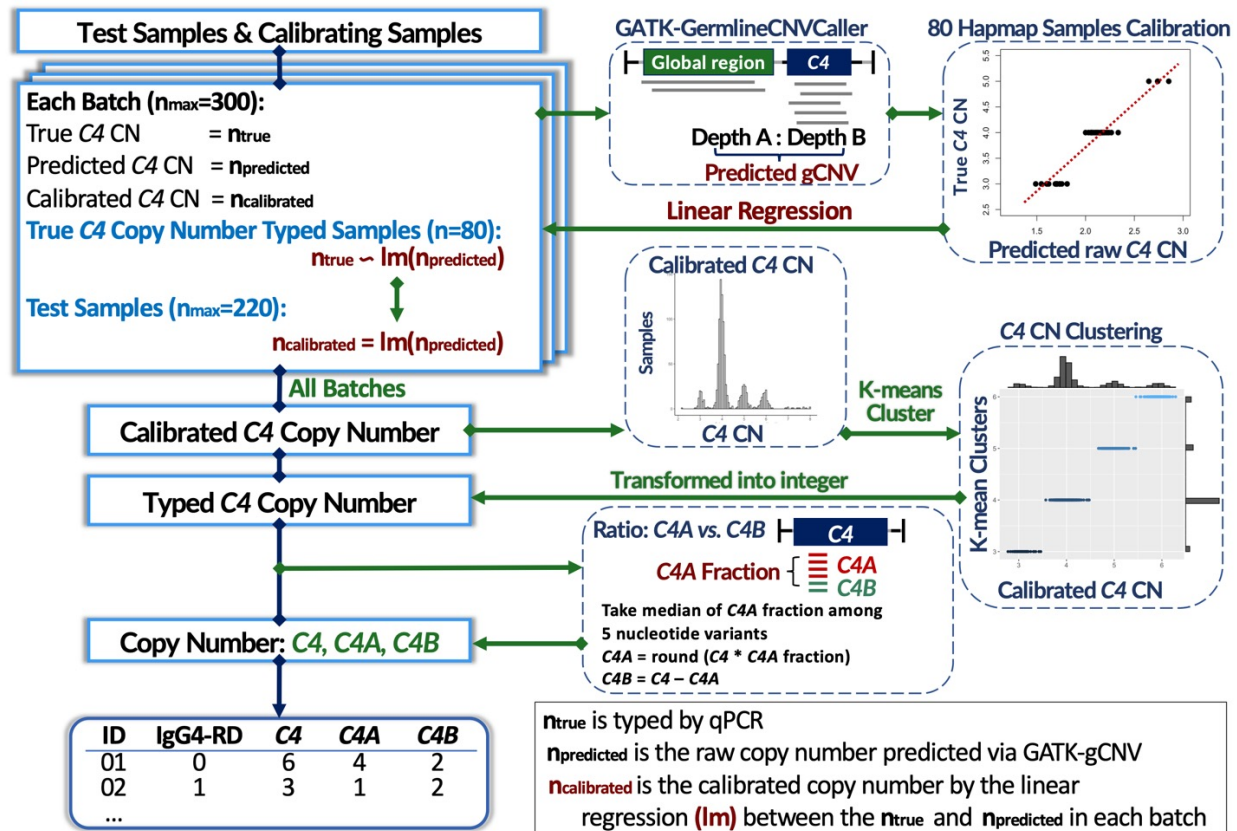
Supplementary Figure 3. The principal component analyses of the study subjects.

(A) Top two principal component (PC) scores of the study samples projected onto the PCs calculated from genome scan data of Biobank Japan. (B) PCA plot based on Set 1 samples, showing individuals that passed quality control. (C) PCA plot of based on Set 2, showing individuals that passed quality control.



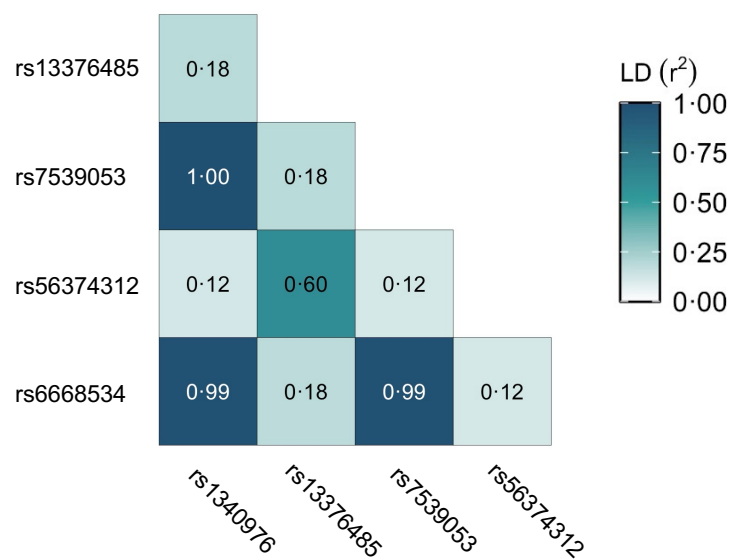
Supplementary Figure 4. CNV estimation workflow of C4 and its two isotypes.

For each batch (maximum $n=300$), test and calibration samples were processed in parallel. Among them, 80 samples with qPCR-calibrated $C4$ copy numbers were used as calibrating samples. Raw $C4$ copy numbers were initially predicted using GATK-GermlineCNVCaller, based on the read depth ratio between the $C4$ gene region and a global reference region. Within the calibrating samples, the predicted copy numbers ($n_{\text{predicted}}$) were regressed on the qPCR-calibrated copy number (n_{true}) to generate a batch-specific linear regression model (lm), which was then applied to case and control samples to obtain calibrated $C4$ copy number ($n_{\text{calibrated}}$). Calibrated $C4$ copy numbers were subsequently clustered using K-means to obtain integer values for the total $C4$ copy number. Copy numbers of $C4A$ and $C4B$ isotypes were then determined using the median $C4A$ fraction derived from five isotype-specific sequences. Final values were calculated as follows: 1) $C4A = \text{round}(C4 \times C4A \text{ fraction})$, 2) $C4B = C4 - C4A$.



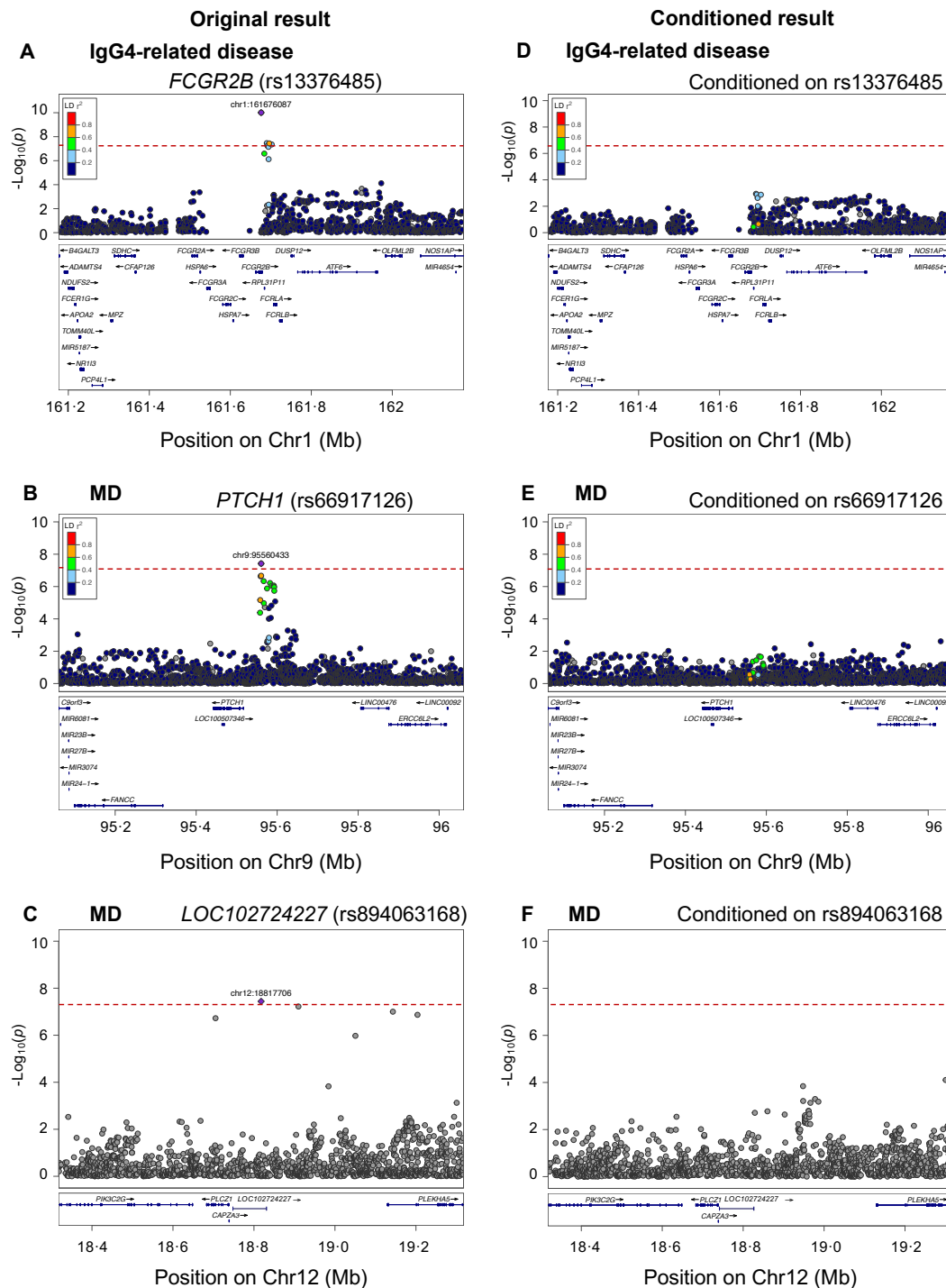
Supplementary Figure 5. Linkage disequilibrium of disease-associated SNVs in the *FCGR2B* locus.

LD plot of four SNVs in the *FCGR2B* locus associated with IgG4-related disease. Pairwise LD(r^2) score is shown in each square with a colour according to the accompanying scale bar.



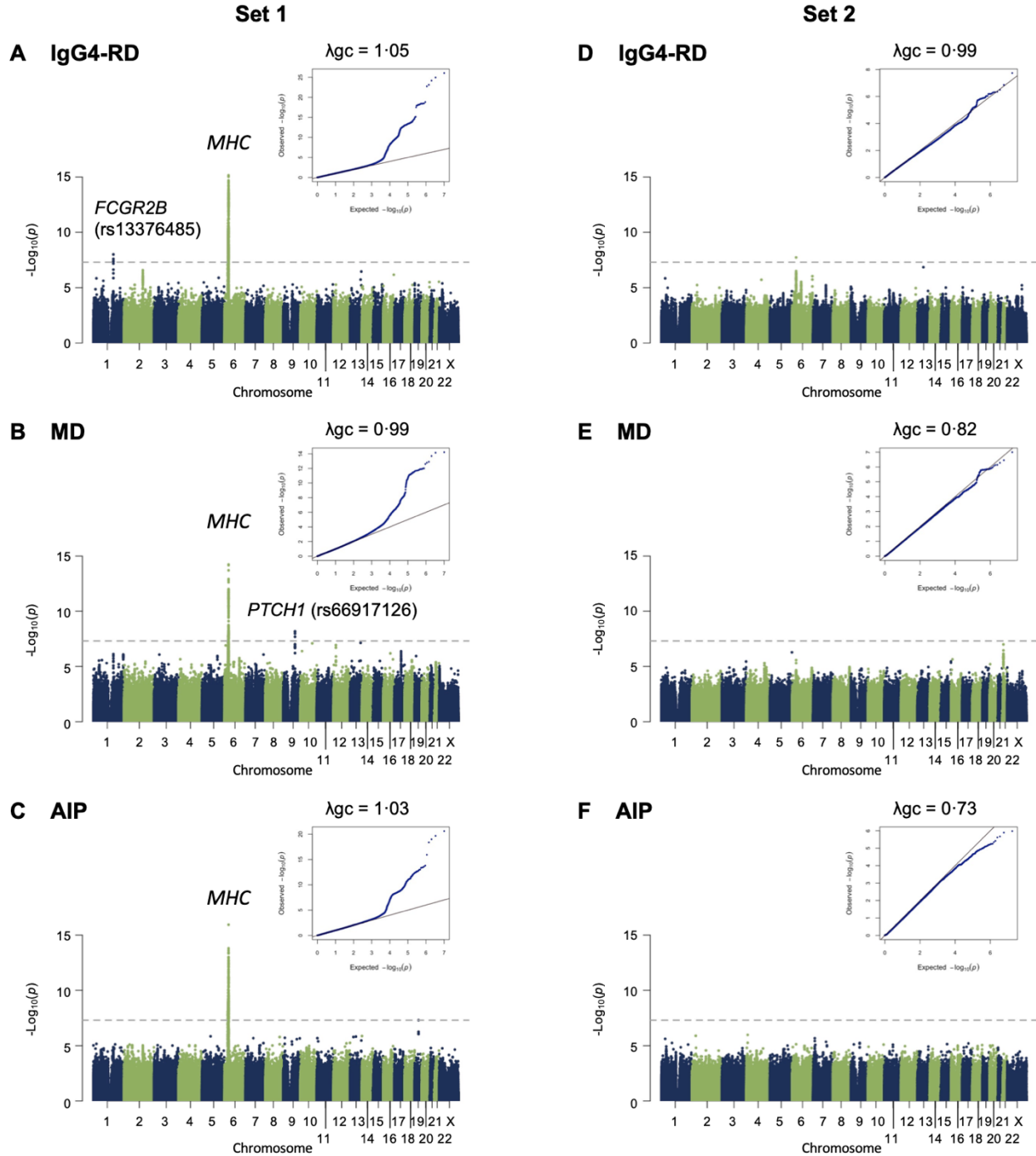
Supplementary Figure 6. Regional plot of chromosomal regions encompassing the associated SNVs.

Regional association plots encompassing SNVs with significant associations. (A) *FCGR2B* and IgG4-related disease, (B) *PTCH1* and MD, and (C) *LOC102724227* and MD. (D), (E) and (F) show the corresponding plots conditioned on the lead SNV at each respective locus. LD score for each SNV with the lead variant is indicated with a colour according to the accompanying scale bar. The dotted horizontal line in the plot represents the significant threshold ($p = 5 \cdot 0 \times 10^{-8}$).



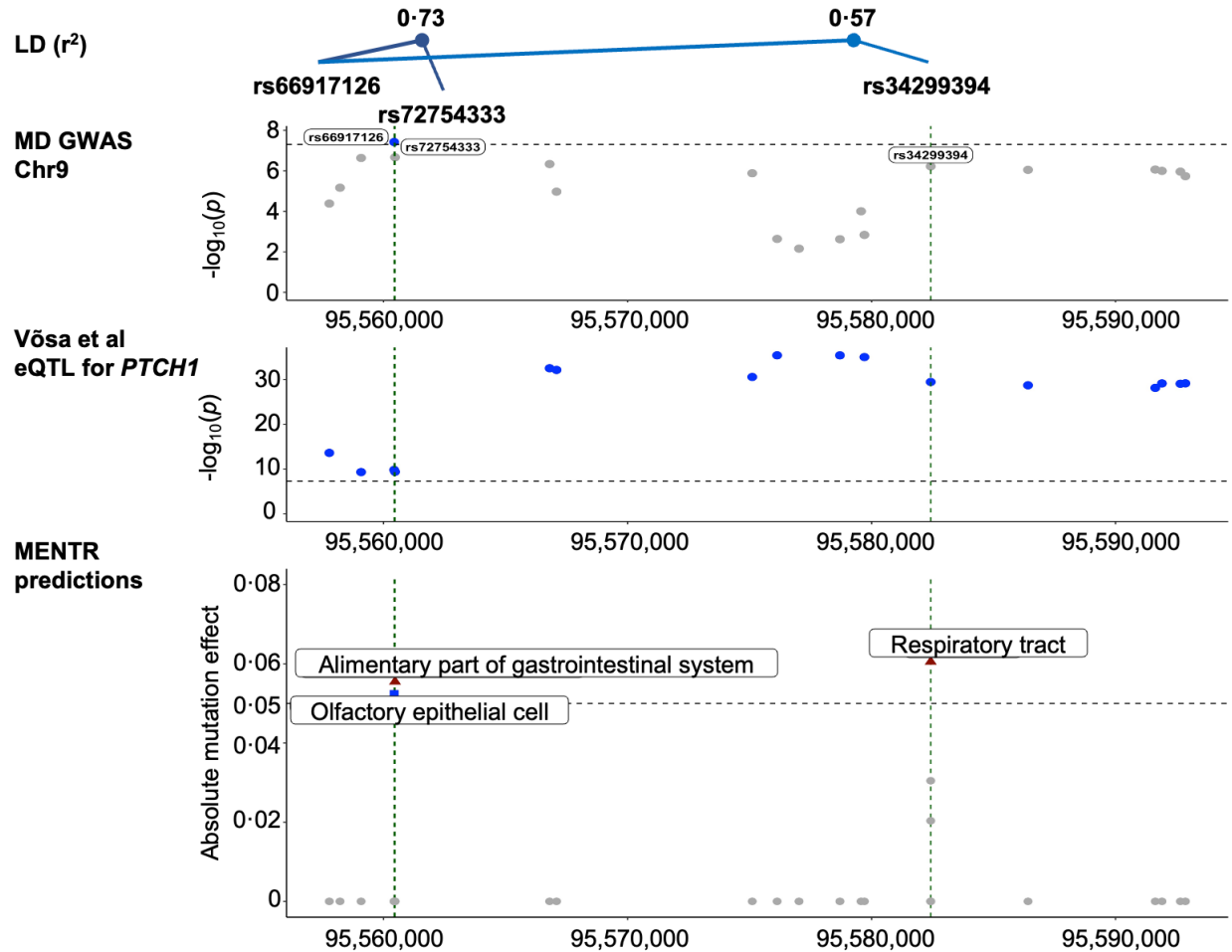
Supplementary Figure 7. GWAS results of IgG4-related disease and its subtypes in sample sets Set 1 and Set 2.

Manhattan plots of (A) Overall individuals in Set 1, (B) Mikulicz's disease and controls in Set 1, (C) autoimmune pancreatitis and controls in Set 1, (D) Overall individuals in Set 2, (E) Mikulicz's disease and controls in Set 2, and (F) autoimmune pancreatitis and controls in Set 2. The dotted horizontal line in the plot represents the significant threshold ($p = 5.0 \times 10^{-8}$). The Q-Q plot is shown in each study with a λ_{gc} score.



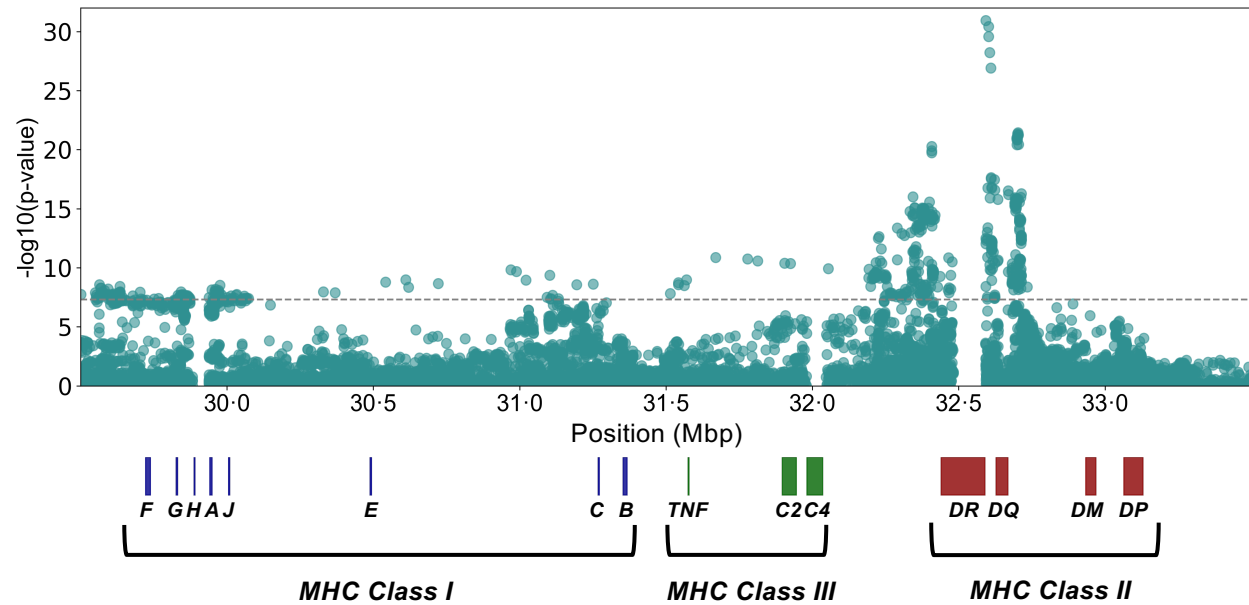
Supplementary Figure 8. Regulatory effects of SNVs associated with MD, eQTL, and MENTR.

Linkage disequilibrium of SNVs, MD meta GWAS result, eQTL results, and mutation effects predicted by MENTR around *PTCH1* were shown. The black horizontal dashed line represents the genome-wide significance threshold ($p=5.0\times10^{-8}$) for MD GWAS and eQTL and the permissive threshold (0.05) for MENTR result.



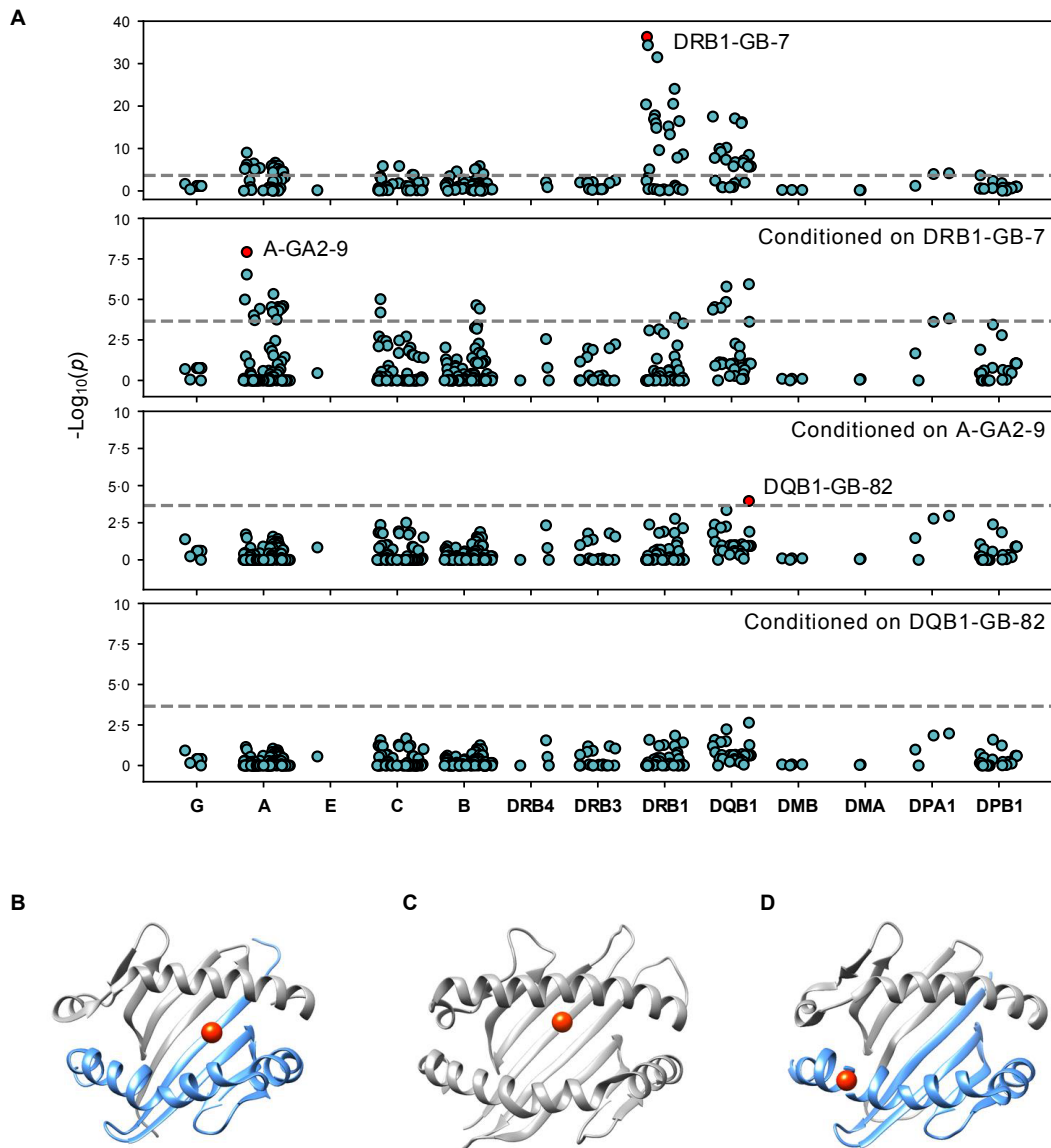
Supplementary Figure 9. Regional plot of GWAS results in the *MHC* locus.

Regional association plot of the *MHC* locus with IgG4-related disease. The x-axis represents chromosomal positions with the location of *HLA* and other genes, and the y-axis shows the $-\log_{10}(p\text{-value})$ from association tests. The horizontal dashed line represents the genome-wide significance threshold ($p=5.0\times 10^{-8}$)



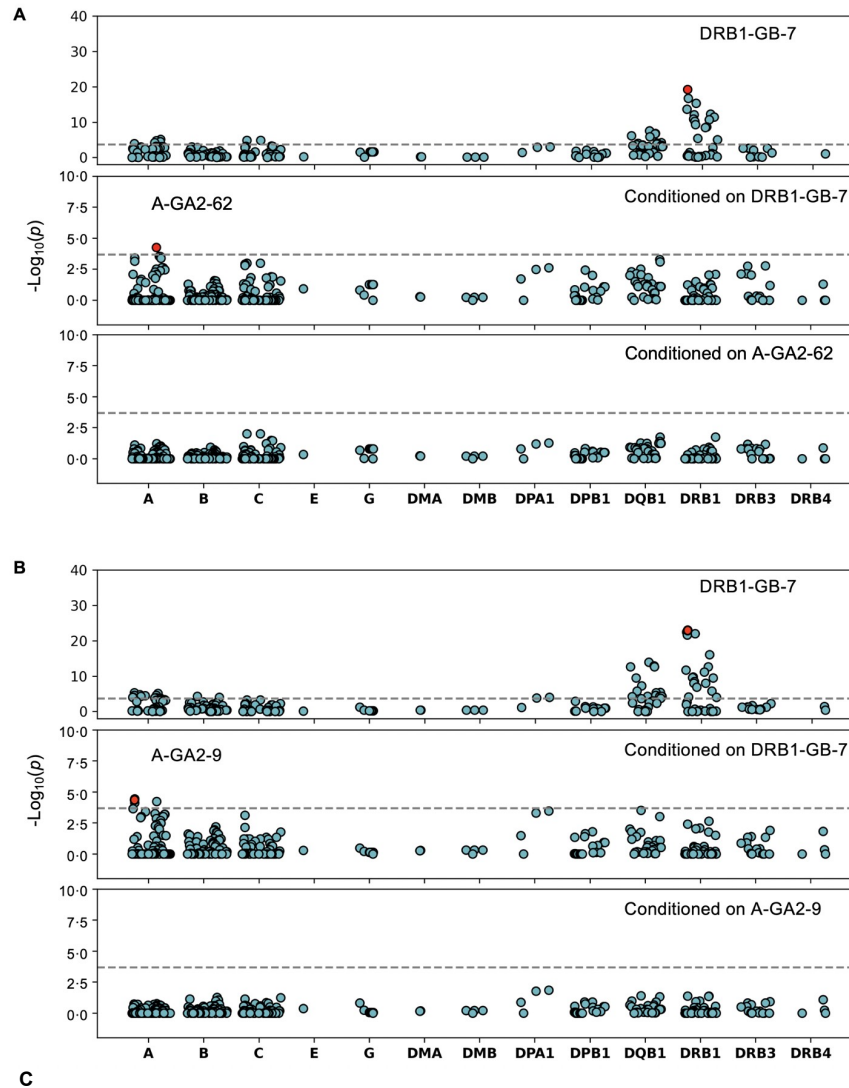
Supplementary Figure 10. Association analysis of HLA amino acid residues with IgG4-related disease

(A) Association between amino acid variations with IgG4-related disease for 228 positions in the peptide-binding groove domain of 13 HLA proteins. The top panel shows the primary association results, while the lower panels present results after conditional analyses. The x-axis represents amino acid position in the peptide-binding groove domain of each protein, and the y-axis shows the $-\log_{10}(p\text{-value})$ from association tests. The horizontal grey dashed line indicates the significance threshold ($p=2.2 \times 10^{-4}$). (B)–(D) Three-dimensional structures of the peptide-binding groove domain of HLA class I and class II proteins, with associated amino acid residues highlighted as red spheres. The α -chain is shown in light grey and the β -chain in light blue. α -helices and β -sheets are depicted as helices and flat ribbons, respectively. (B) DRB1-GB-7 is located at the centre of the G β domain of the HLA-DRB1 protein. (C) A-GA2-9 is located at the centre of the G α 2 domain of the HLA-A protein. (D) DQB1-GB-82 is located on the α -helix in the G β domain of the HLA-DQB1 protein.



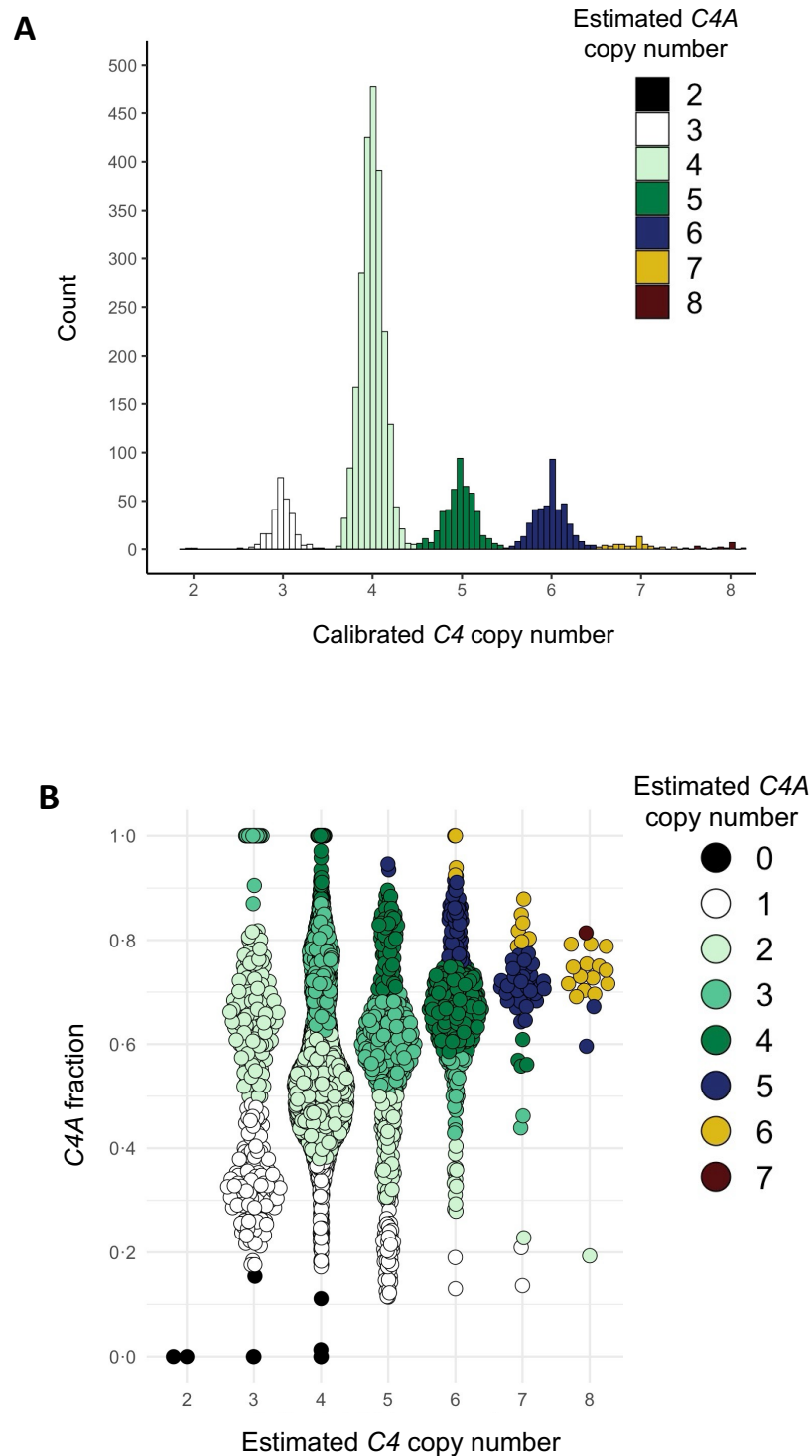
Supplementary Figure 11. Association analysis of HLA amino acid residues with IgG4-related disease subtypes.

Association between amino acid variations with (A) MD and (B) AIP for 228 positions in the peptide-binding groove domain of 13 HLA proteins. The top panel shows the primary association results, while the lower panels present results after conditional analyses. The x-axis represents amino acid position in the peptide-binding groove domain of each protein, and the y-axis shows the $-\log_{10}(p\text{-value})$ from association tests. The horizontal grey dashed line indicates the significance threshold ($p=2.2 \times 10^{-4}$). (C) The position of A-GA2-62 (in red sphere) in the three-dimensional structure of the HLA-A peptide-binding groove domain.



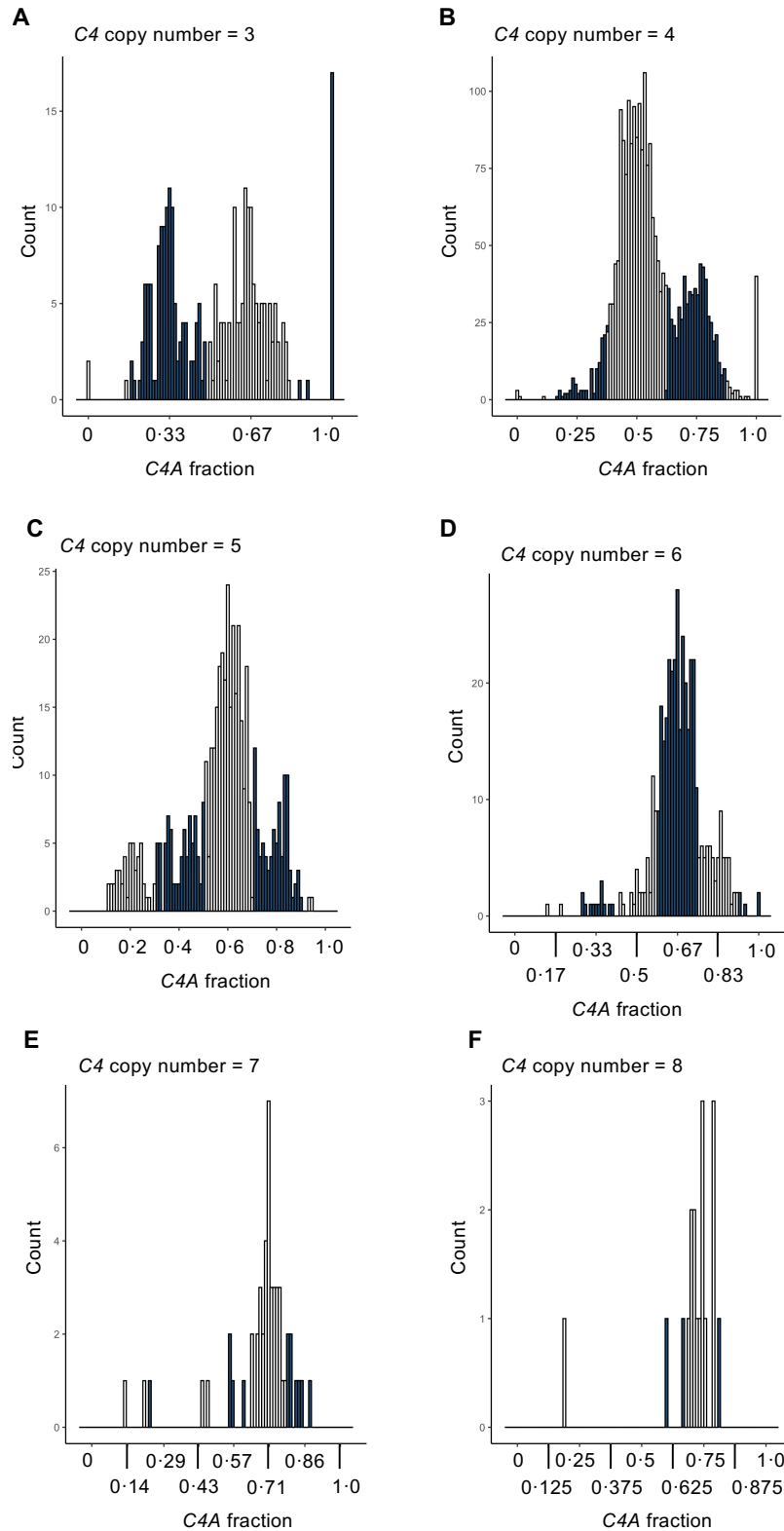
Supplementary Figure 12. Distribution of *C4* copy number and estimated fraction of *C4A* in each group.

(A) Distribution of estimated *C4* copy numbers in all individuals recruited in this study (n=3,528). (B) The fraction of *C4A* in the seven groups of individuals with an estimated *C4* copy number from two to eight. Each circle represents an individual.



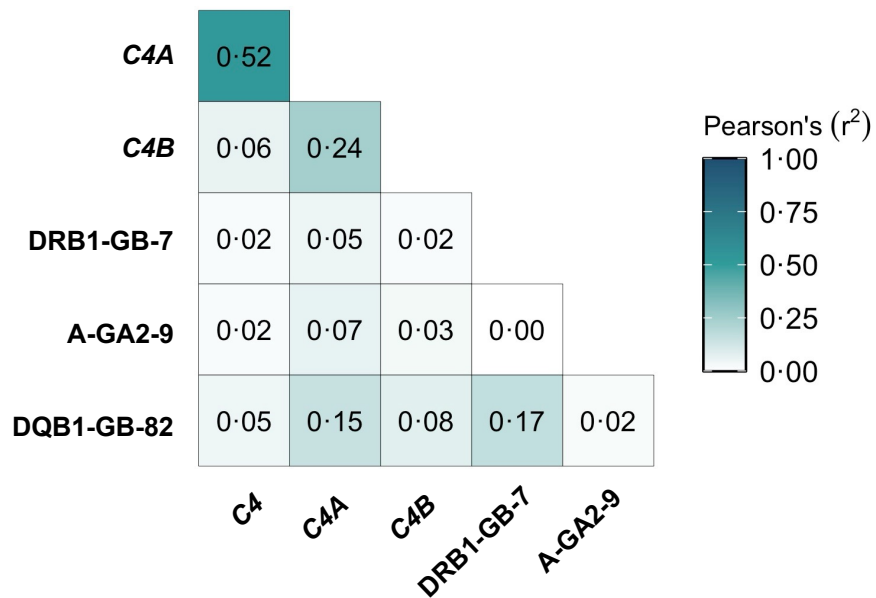
Supplementary Figure 13. Distribution of *C4A* fractions stratified by *C4* copy number.

Each histogram shows the distribution of *C4A* copy numbers in subpopulations with total *C4* copy numbers of (A) three, (B) four, (C) five, (D) six, (E) seven, and (F) eight.



Supplementary Figure 14. Correlation between *C4* CNVs and HLA amino acid residues.

Correlation plot (Pearson's r^2) between *C4*, *C4A* and *C4B* copy numbers and identified amino acid positions associated with IgG4-related disease. Pairwise Pearson's r^2 score is shown in each square with a colour according to the accompanying scale bar.



Supplementary Table 1. The list of the participating institutes in the Japanese IgG4-Related Disease Working Consortium.

Institute	Department	Study set
Aichi Cancer Center Hospital	Department of Gastroenterology	Set 1
Fujita Health University Bantane Hospital	Department of Gastroenterology	Set 1
Hyogo Prefectural Amagasaki Hospital	Department of Gastroenterology	Set 1
Japan Community Health Care Organization Isahaya General Hospital	Department of Rheumatology	Set 1
Japanese Red Cross Osaka Hospital	Department of Gastroenterology	Set 1 and 2
Japanese Red Cross Otsu Hospital	Department of Gastroenterology	Set 1
Japanese Red Cross Wakayama Medical Center	Department of Gastroenterology	Set 1
Jichi Medical University	Division of Gastroenterology, Department of Medicine	Set 2
Kagawa University Faculty of Medicine	Division of Hematology, Rheumatology and Respiratory Medicine, Department of Internal Medicine	Set 2
Kagoshima University Graduate School of Medical and Dental Sciences	Digestive and Lifestyle Diseases	Set 1
Kanazawa Medical University	Department of Hematology and Immunology	Set 1
Kanazawa University Hospital	Department of Rheumatology	Set 1
Kansai Electric Power Hospital	Department of Gastroenterology and Hepatology	Set 1
Kansai Medical University	The Third Department of Internal Medicine	Set 1
Kansai Medical University Hirakata Hospital	Department of Gastroenterology and Hepatology	Set 1
Keio University School of Medicine	Department of Ophthalmology	Set 1
Keio University School of Medicine	Division of Gastroenterology and Hepatology, Department of Internal Medicine	Set 1
Kishiwada City Hospital	Department of Gastroenterology	Set 1
Kobe City Medical Center General Hospital	Department of Gastroenterology	Set 1
Kobe City Nishi-Kobe Medical Center	Department of Gastroenterology	Set 1
Kobe Kaisei Hospital	Department of Ophthalmology	Set 1
Kobe University Hospital	Department of Gastroenterology	Set 1
Kyoto University Hospital	Department of Gastroenterology and Hepatology	Set 1
Kyoto University Hospital	Department of Rheumatology and Clinical Immunology	Set 1
Kyoto-Katsura Hospital	Center for Clinical Study	Set 1
Kyushu University Hospital	Department of Oral Surgery	Set 1 and 2

Medical Research Institute Kitano Hospital, PIIF Tazuke-Kofukai	Digestive Disease Center	Set 1
Nagano Municipal Hospital	Department of Gastroenterology	Set 1
Nagaoka Red Cross Hospital	Department of Internal Medicine	Set 1
Nagasaki University Graduate School of Biomedical Sciences	Unit of Rehabilitation Sciences	Set 1
Nagoya City University Graduate School of Medical Sciences	Department of Gastroenterology and Metabolism	Set 1 and 2
Nagoya University Graduate School of Medicine	Department of Pathology and Laboratory Medicine/Diagnostic Pathology	Set 1
National Hospital Organization Kyoto Medical Center	Clinical Research Center	Set 1
National Hospital Organization Okayama Medical Center	Department of Ophthalmology	Set 1
Saiseikai Ibaraki Hospital	Department of Gastroenterology	Set 1
Sapporo Medical University	Department of Gastroenterology, Rheumatology and Clinical Immunology	Set 1 and 2
Shiga General Hospital	Department of Gastroenterology	Set 1
Shimane University Hospital	Department of Gastroenterology and Hepatology	Set 1
Shinshu University	Center for Health Safety and Environmental Management	Set 1
Takamatsu Red Cross Hospital	Department of Gastroenterology	Set 1
Takatsuki Red Cross Hospital	Department of Gastroenterology	Set 1
Teine Keijinkai Hospital	Department of Gastroenterology	Set 1
Tenri Hospital	Department of Gastroenterology	Set 1
The University of Tokyo Hospital	Department of Gastroenterology	Set 1
The University of Toyama	Center for Health Care and Human Sciences	Set 1 and 2
Tohoku University Hospital	Department of Gastroenterology	Set 1 and 2
Tokai University Hachioji Hospital	Department of Gastroenterology	Set 1
Tokyo Metropolitan Cancer and Infectious Diseases Center Komagome Hospital	Department of Internal Medicine	Set 1
Tokyo Woman's Medical University Yachiyo Medical Center	Department of Gastroenterology	Set 1
University of Occupational and Environmental Health	The First Department of Internal Medicine	Set 1
Wakayama Medical University	The 1st Department of Medicine	Set 1
Yokohama City University Hospital	Department of Gastroenterology	Set 1

Supplementary Table 2. Summary of the whole-genome sequencing results.

	Set 1				Set 2	
	Case		Control		Case	Control
Number of subjects	703		3,148		231	419
Sequence platform	NovaSeq		HiSeqX		NovaSeq	HiSeqX
Target depth (number of subjects)	15× (131)		30× (572)	15×	30×	30×
Mean read length (bp)	100 (63)	150 (68)	150	150	150	150
Mean of base count	55,970,800,000	54,780,600,000	95,045,300,000	55,145,700,000	106,793,000,000	129,594,000,000
Read length (mean±SD)	99.9 ± 2.6	150.3 ± 8.5	149.3 ± 8.6	150.1 ± 7.5	149.2 ± 9.1	148.7 ± 13.1
Average coverage (mean±SD)	17.4 ± 1.1	17.0 ± 2.5	30.1 ± 0.9	17.8 ± 3.5	31 ± 1.7	41.8 ± 1.6

SD=standard deviation.

Supplementary Table 3. Summary of SNVs significantly associated with IgG4-related disease in Set 1, Set 2, and the meta-analysis.

rsID	Chr.	Position (hg38)	Allele (Ref./Alt.)	Closest gene	Freq. Alt. allele in jMorp	Set	Freq. Alt. allele			<i>p</i> -value	OR	95% CI	Direction [†]	I ²
							All	Case	Control					
IgG4-related disease														
rs13376485	1	161,676,087	G/A	FCGR2B	0.232	1	0.214	0.146	0.233	9.8×10 ⁻⁹	0.59	0.49-0.71	--	0.00
						2	0.199	0.150	0.226	0.0027	0.61	0.44-0.84		
						Meta	0.211	0.147	0.232	9.8×10 ⁻¹¹	0.59	0.51-0.69		
rs7539053	1	161,690,055	T/C	FCGR2B	0.619	1	0.592	0.518	0.614	2.5×10 ⁻⁸	0.68	0.59-0.78	--	0.59
						2	0.575	0.547	0.590	0.18	0.85	0.66-1.08		
						Meta	0.589	0.525	0.610	3.4×10 ⁻⁸	0.71	0.63-0.80		
rs56374312	1	161,695,965	C/T	FCGR2B	0.155	1	0.142	0.092	0.157	2.3×10 ⁻⁷	0.55	0.44-0.69	--	0.00
						2	0.134	0.105	0.149	0.038	0.67	0.46-0.98		
						Meta	0.141	0.095	0.156	3.6×10 ⁻⁸	0.58	0.48-0.70		
rs6668534	1	161,702,205	G/A	FCRLA	0.619	1	0.594	0.520	0.615	3.9×10 ⁻⁸	0.68	0.59-0.78	--	0.54
						2	0.576	0.547	0.592	0.16	0.84	0.66-1.07		
						Meta	0.590	0.527	0.611	4.4×10 ⁻⁸	0.72	0.64-0.81		
MD														
rs66917126	9	95,560,433	C/T	PTCHI	0.102	1	0.100	0.167	0.092	1.1×10 ⁻⁸	2.17	1.66-2.82	++	0.66
						2	0.114	0.137	0.109	0.38	1.27	0.75-2.14		
						Meta	0.102	0.159	0.094	3.8×10 ⁻⁸	1.94	1.53-2.46		
rs894063168	12	18,817,706	C/T	LOC102724 227	0.003	1	0.004	0.021	0.002	1.2×10 ⁻⁷	13.92	5.25-36.86	++	0.00
						2	0.004	0.011	0.002	0.11	6.89	0.65-73.33		
						Meta	0.004	0.018	0.002	3.6×10 ⁻⁸	12.57	5.11-30.93		

Chr.=chromosome; Ref.=reference; Alt.=alternative; Freq.=frequency; OR=odds ratio; CI=confidence interval; I²=percentage of variation due to heterogeneity.

jMorp: Japanese Multi-Omics Reference Panel (Tadaka *et al. Nucleic Acids Research*, 2023, doi: 10.1093/nar/gkad978)

[†]--: associations with odds ratios <1.00 in all sets, ++: associations with odds ratios >1.00 in all sets.

Supplementary Table 4. Credible SNVs by fine-mapping analysis.

Cohort	Lead variant	Variants in credible set				PIP*	Alt. allele frequency			<i>p</i> -value
		rsID	Chr.	Position (hg38)	Allele (Ref./Alt.)		All	Case	Ctrl	
IgG4-related disease (Case n=869; Ctrl n=2,659)										
	rs13376485	rs13376485	1	161,676,087	G/A	0.978	0.211	0.147	0.232	9.8×10 ⁻¹¹
MD (Case n=363; Ctrl n=2,659)										
	rs66917126	rs66917126	9	95,560,433	C/T	0.495	0.102	0.159	0.094	3.8×10 ⁻⁸
	rs66917126	rs72754333	9	95,560,469	G/T	0.110	0.132	0.192	0.124	2.2×10 ⁻⁷
	rs66917126	rs72754331	9	95,559,079	G/A	0.102	0.135	0.193	0.127	2.3×10 ⁻⁷
	rs66917126	rs28675274	9	95,566,806	C/T	0.056	0.148	0.204	0.140	4.6×10 ⁻⁷
	rs66917126	rs34299394	9	95,582,424	A/T	0.045	0.137	0.192	0.130	6.0×10 ⁻⁷
	rs66917126	rs2405060	9	95,591,628	A/C	0.033	0.150	0.206	0.143	8.7×10 ⁻⁷
	rs66917126	rs34777564	9	95,586,412	G/A	0.032	0.138	0.191	0.130	9.0×10 ⁻⁷
	rs66917126	rs4335262	9	95,591,902	A/G	0.029	0.151	0.207	0.144	1.0×10 ⁻⁶
	rs66917126	rs10118174	9	95,592,658	T/C	0.027	0.151	0.206	0.143	1.1×10 ⁻⁶
	rs66917126	rs55957307	9	95,575,110	T/C	0.023	0.143	0.194	0.136	1.3×10 ⁻⁶
	rs894063168	rs894063168	12	18,817,706	C/T	0.374	0.004	0.018	0.002	3.6×10 ⁻⁸
	rs894063168	rs1324549138	12	18,909,843	G/C	0.241	0.004	0.018	0.002	6.0×10 ⁻⁸
	rs894063168	rs1231784939	12	19,143,015	T/A	0.155	0.004	0.018	0.002	1.0×10 ⁻⁷
	rs894063168	N/A	12	19,203,824	T/TAATA	0.119	0.004	0.018	0.002	1.4×10 ⁻⁷
	rs894063168	rs149235894	12	18,705,115	T/G	0.088	0.010	0.029	0.007	1.9×10 ⁻⁷

Chr.=chromosome; Ref.=reference; Alt.=alternative; PIP=posterior inclusion probability; MD=Mikulicz's disease.

*Variants are listed in descending PIP until the cumulative PIP first exceeds 95%.

Supplementary Table 5. Heterogeneity of newly detected SNVs in different IgG4-related disease subsets.

Cohort	rsID	Chr.	Position (hg38)	Allele (Ref./Alt.)	IgG4-related disease vs. MD			MD vs. AIP		
					Q	I ² (%)	<i>p</i> -value	Q	I ² (%)	<i>p</i> -value
IgG4-related disease (Case n=869, Control n=2659)										
	rs13376485	1	161,676,087	G/A	0.475	0.0	0.49	1.740	42.5	0.19
	rs7539053	1	161,690,055	T/C	0.684	0.0	0.41	0.762	0.0	0.38
	rs56374312	1	161,695,965	C/T	0.924	0.0	0.34	2.550	60.8	0.11
	rs6668534	1	161,702,205	G/A	0.546	0.0	0.46	0.671	0.0	0.41
MD (Case n=363, Control n=2659)										
	rs66917126	9	95,560,433	C/T	7.058	85.8	0.0079	9.463	89.4	0.0021
	rs894063168	12	18,817,706	C/T	0.741	0.0	0.39	1.219	18.0	0.27

Chr.=chromosome; Ref.=reference; Alt.=alternative; Q=Cochran's Q statistic; I²=percentage of variation due to heterogeneity; MD=Mikulicz's disease; AIP=autoimmune pancreatitis.

Supplementary Table 6. eQTL results for the variants in Linkage Disequilibrium with rs66917126 in the *PTCH* region.

rsID	Chr.	Position (hg38)	Allele (Ref./Alt.)	LD (r^2) with rs66917126*	GWAS (MD)			eQTL [†]	
					OR	95% CI	<i>p</i> -value	Effect size	<i>p</i> -value
rs77880503	9	95557776	G/A	0.64	1.66	1.30 - 2.11	4.1×10^{-5}	0.151	2.5×10^{-14}
rs72754331	9	95559079	G/A	0.71	1.77	1.43 - 2.20	2.3×10^{-7}	0.101	4.9×10^{-10}
rs66917126	9	95560433	C/T	1.00	1.94	1.53 - 2.46	3.8×10^{-8}	0.105	1.6×10^{-10}
rs72754333	9	95560469	G/T	0.73	1.78	1.43 - 2.21	2.2×10^{-7}	0.102	4.2×10^{-10}
rs28675274	9	95566806	C/T	0.47	1.72	1.39 - 2.13	4.6×10^{-7}	0.244	3.0×10^{-33}
rs76099988	9	95567089	T/A	0.56	1.61	1.30 - 1.99	1.1×10^{-5}	0.250	7.0×10^{-33}
rs55957307	9	95575110	T/C	0.54	1.70	1.37 - 2.10	1.3×10^{-6}	0.239	2.6×10^{-31}
rs79934023	9	95576130	G/A	0.35	1.47	1.15 - 1.88	2.3×10^{-4}	0.269	3.6×10^{-36}
rs78621370	9	95578703	C/T	0.35	1.46	1.14 - 1.87	2.4×10^{-3}	0.269	3.9×10^{-36}
rs2897052	9	95579706	G/C	0.34	1.48	1.16 - 1.89	1.5×10^{-3}	0.267	9.4×10^{-36}
rs34299394	9	95582424	A/T	0.57	1.74	1.40 - 2.16	6.0×10^{-7}	0.226	3.4×10^{-30}
rs34777564	9	95586412	G/A	0.57	1.72	1.39 - 2.14	9.0×10^{-7}	0.224	1.9×10^{-29}
rs2405060	9	95591628	A/C	0.50	1.69	1.37 - 2.08	8.7×10^{-7}	0.221	7.2×10^{-29}
rs4335262	9	95591902	A/G	0.51	1.68	1.37 - 2.08	1.0×10^{-6}	0.223	7.2×10^{-30}
rs10118174	9	95592658	T/C	0.51	1.68	1.36 - 2.07	1.1×10^{-6}	0.223	8.7×10^{-30}
rs10981328	9	95592863	C/T	0.51	1.66	1.35 - 2.05	1.8×10^{-6}	0.223	6.9×10^{-30}

Chr.=chromosome; Ref.=reference; Alt.=alternative; LD=Linkage Disequilibrium; MD=Mikulicz's disease; OR=odds ratio; CI=confidence interval.

*LD was estimated from the control samples of Set 1 and Set 2.

[†]Vösa U *et al. Nat Genet* 2021; **53**: 1300–10.

Supplementary Table 7. Predicted effect of variants associated with IgG4-related disease.

rsID	Chr.	Position (hg19)	Position (hg38)	Allele (Ref./Alt.)	LD (r ²) with rs66917126*	Alt. allele frequency			Distance to TSS (bp)	p-value
						MD	AIP	Control		
rs66917126	9	98,322,715	95,560,433	C/T	1.00	0.159	0.101	0.094	43,508	3.8×10 ⁻⁸
rs72754333	9	98,322,751	95,560,469	G/T	0.74	0.192	0.132	0.124	7,042	2.2×10 ⁻⁷
rs34299394	9	98,344,706	95,582,424	A/T	0.58	0.192	0.130	0.130	150	6.0×10 ⁻⁷

Chr.=chromosome; Ref.=reference; Alt.=alternative; LD=linkage disequilibrium; MD=Mikulicz's disease; AIP=autoimmune pancreatitis; TSS=transcription start site; N.A.=not available.

*LD was estimated using pooled control samples of Set 1 and Set 2.

rsID	CAGE ID (hg19)/Strand	Type	Name	Gene Class	Predicted effect
rs66917126	chr9:98279198..98279213/–	Promoter	<i>PTCH1</i>	coding_mRNA	Olfactory epithelial cell: –0.052
rs72754333	chr9:98315655..98315762/N	Enhancer	ADDG09098315655.E	enhancer	Alimentary part of gastrointestinal system: 0.056
rs34299394	chr9:98344805..98344906/N	Enhancer	ADDG09098344805.E	enhancer	Respiratory tract: 0.061

Supplementary Table 8. Association of rs13376485 with IgG4-related disease, MD, and AIP.

Cohort	Frequency of A allele		OR	95% CI	<i>p</i> -value
	Case	Control			
IgG4-related disease (n=869)	0·147	0·232	0·59	0·51-0·69	$9·8 \times 10^{-11}$
MD (n=363)	0·132	0·232	0·54	0·42-0·68	$1·7 \times 10^{-7}$
AIP (n=457)	0·158	0·232	0·66	0·54-0·80	$4·2 \times 10^{-5}$

OR=odds ratio; CI=confidence interval; MD=Mikulicz's disease; AIP=autoimmune pancreatitis.

Supplementary Table 9. HLA amino acid residues associated with IgG4-related disease subtypes.

Amino acid residue	Amino acid frequency			<i>p</i> -value*	OR	95% CI	Corresponding <i>HLA</i> alleles [†]
	Overall	Case	Control				
MD (Case n=363; Control n=2,659)							
DRB1-GB-7							
Ser	0.320	0.351	0.377	Reference	-	-	*08:02, *08:03, *11:01, *12:01, *12:02, *13:02, *14:03, *14:05, *14:06, *14:54
Val	0.211	0.376	0.228	3.2×10⁻⁹	1.74	1.45-2.09	*04:03, *04:05, *04:06, *04:10
Asp	0.110	0.051	0.139	4.8×10⁻⁷	0.40	0.28-0.58	*09:01
Leu	0.056	0.056	0.067	0.63	0.92	0.65-1.30	*01:01
Pro	0.159	0.165	0.189	0.59	0.94	0.74-1.19	*15:01, *15:02
A-GA2-62							
His	0.697	0.766	0.821	Reference	-	-	*02:01, *02:06, *02:07, *11:01, *24:02, *26:01, *26:02, *26:03
Arg	0.159	0.234	0.179	4.7×10 ⁻⁴	1.39	1.15-1.66	*31:01, *33:03
AIP (Case n=457; Control n=2,659)							
DRB1-GB-7							
Ser	0.326	0.324	0.377	Reference	-	-	*08:02, *08:03, *11:01, *12:01, *12:02, *13:02, *14:03, *14:05, *14:06, *14:54
Val	0.222	0.389	0.228	1.3×10⁻¹⁴	1.94	1.64-2.29	*04:03, *04:05, *04:06, *04:10
Asp	0.114	0.067	0.139	6.9×10⁻⁵	0.56	0.42-0.75	*09:01
Leu	0.057	0.054	0.067	0.65	0.93	0.67-1.28	*01:01
Pro	0.164	0.166	0.189	0.95	1.01	0.81-1.25	*15:01, *15:02
A-GA2-9							
Tyr	0.523	0.659	0.581	Reference	-	-	*02:01, *02:06, *26:01, *26:02, *26:03, *31:01, *33:03
Phe	0.330	0.324	0.382	1.6×10 ⁻⁴	0.75	0.64-0.87	*24:02
Cys	0.030	0.018	0.036	0.015	0.43	0.26-0.72	*02:07, *11:01

OR=odds ratio; MD=Mikulicz's disease; AIP=autoimmune pancreatitis.

*Statistically significant values with $p < 1.53 \times 10^{-4}$ are shown in bold.

†*HLA* alleles with a frequency of greater than 0.01 in either the case and control are shown.

Supplementary Table 10. Heterogeneity of IgG4-related disease associated HLA amino acid residues in Set 1 and Set 2.

Amino acid residue	Set	Case		Control		Effect size	SE	OR	95%CI	Q	I ²
		Frequency	Count	Frequency	Count						
DRB1-GB-7-Val	Set 1	0.381	492	0.225	1,013	0.752	0.069	2.12	1.85-2.43	0.161	0
	Set 2	0.368	164	0.247	200	0.573	0.132	1.77	1.37-2.30		
A-GA2-9-Phe	Set 1	0.335	433	0.191	859	-0.197	0.068	0.82	0.72-0.94	0.173	0
	Set 2	0.307	137	0.381	309	-0.382	0.129	0.68	0.53-0.88		
DQB1-GB-82-Ala	Set 1	0.332	429	0.405	1,827	-0.315	0.068	0.73	0.64-0.83	0.051	0
	Set 2	0.285	127	0.377	305	-0.416	0.132	0.66	0.51-0.85		

SE=standard error; OR=odds ratio; Q=Cochran's Q statistic; I²=percentage of variation due to heterogeneity.

Supplementary Table 11. Results of association analyses of IgG4-related disease risk against *C4* copy number variation.

Model	<i>C4</i>			<i>C4A</i>			<i>C4B</i>			Δ AIC
	beta	SE	<i>p</i> -value	beta	SE	<i>p</i> -value	beta	SE	<i>p</i> -value	
<i>C4</i>	− 0.068	0.045	0.14							21.01
<i>C4A</i>				− 0.18	0.042	1.2×10^{-5}				3.44
<i>C4B</i>							0.24	0.056	2.2×10^{-5}	5.25
<i>C4A</i> + <i>C4B</i>				− 0.13	0.048	0.0079	0.15	0.064	0.019	0 (best)

SE: standard error, AIC: Akaike information criterion.

Supplementary Table 12. Predicted relative onset risk of IgG4-related disease based on *C4* CNV combinations.

Copy number*		Odds	SE	OR to reference	95%CI	IgG4-related disease (n=869)		Control (n=2,659)	
<i>C4A</i>	<i>C4B</i>					Count	Frequency	Count	Frequency
1	4	0.55	0.126	1.54	1.18-1.99	15	0.017	32	0.012
2	4	0.48	0.134	1.35	1.03-1.78	5	0.006	8	0.003
1	3	0.47	0.082	1.32	1.10-1.58	28	0.032	84	0.032
2	3	0.42	0.076	1.16	0.98-1.38	21	0.024	60	0.023
1	2	0.41	0.077	1.14	0.96-1.35	38	0.044	81	0.031
3	3	0.37	0.097	1.02	0.83-1.26	7	0.008	37	0.014
2	2	0.36	0.044	Reference	-	424	0.488	1,069	0.402
3	2	0.32	0.050	0.88	0.77-1.00	61	0.070	194	0.073
2	1	0.31	0.080	0.86	0.72-1.03	25	0.029	104	0.039
4	2	0.28	0.087	0.78	0.64-0.94	64	0.074	219	0.082
3	1	0.27	0.062	0.76	0.65-0.88	121	0.139	502	0.189
5	2	0.24	0.132	0.68	0.52-0.90	3	0.004	31	0.012
4	1	0.24	0.077	0.67	0.56-0.79	23	0.027	68	0.026
3	0	0.23	0.116	0.65	0.51-0.83	2	0.002	17	0.006
6	2	0.22	0.178	0.60	0.42-0.86	6	0.007	8	0.003
5	1	0.21	0.112	0.59	0.46-0.74	10	0.012	60	0.023
4	0	0.21	0.112	0.57	0.45-0.73	10	0.012	51	0.019
6	1	0.19	0.154	0.52	0.38-0.71	1	0.001	7	0.003

OR=odds ratio; SE=standard error.

*Combinations with a frequency <0.002 in the control group were excluded due to rarity.

Supplementary Table 13. Multi-variate logistic regression of *C4A* and HLA-DQB1-GB-82.

Model	<i>C4A</i> CNV			DQB1-GB-82-Ala		
	OR	95% CI	<i>p</i> -value	OR	95% CI	<i>p</i> -value
<i>C4A</i> CNV	0·83	0·77-0·90	$1·2 \times 10^{-5}$	-	-	-
HLA-DQB1-GB-82-Ala	-	-	-	0·70	0·62-0·79	$4·7 \times 10^{-9}$
<i>C4A</i> CNV + HLA-DQB1-GB-82-Ala	0·89	0·81-0·97	$9·7 \times 10^{-3}$	0·72	0·63-0·82	$6·8 \times 10^{-7}$

OR=odds ratio; CI=confidence interval.