

Supplementary appendix 1

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

Supplement to: Attfield KE, Armen AP, Balaram Kuttikkatte S, et al. Identification of genetic risk loci associated with aquaporin 4-positive neuromyelitis optica spectrum disorder: a genome-wide association study. *Lancet Neurol* 2026; **25**: 482–91.

Table of Contents

Supplementary tables	4
Supplementary table 1: Centre statistics.	4
Supplementary table 2: Disease onset pattern statistics.	5
Supplementary table 3: Initial attack recovery statistics.	5
Supplementary table 4: Ethnic background statistics.	5
Supplementary table 5: Number of samples per ethnic background and 1KG superpopulation.	5
Supplementary table 6: Association statistics for each study.	5
Supplementary table 7: Matching coefficients of selected autoimmune diseases with AQP4-positive NMOSD.	6
Supplementary table 8: Correlations of selected autoimmune diseases with AQP4-positive NMOSD.	6
Supplementary figures	7
Supplementary figure 1: Flow chart of the process of generating the genome-wide and extended chromosome-6 datasets.	7
Supplementary figure 2: UMAP plot of the samples, coloured by ethnic background.	8
Supplementary figure 3: UMAP plot of the samples, coloured by country of centre of origin.	8
Supplementary figure 4: UMAP plot of the samples, coloured by 1KG superpopulation.	9
Supplementary figure 5: UpSet plot of sample reasons for removal.	10
Supplementary figure 6: QQ plot for the pan-ancestry study.	11
Supplementary figure 7: Histogram of sample distance from the nearest 1KG sample.	12
Supplementary figure 8: QQ plot for the European study.	13
Supplementary figure 9: Graph of genetic correlations.	14
Supplementary figure 10: Heatmap of comparative treatment efficacy across selected CNS, PNS, and systemic autoimmune diseases.	15
Supplementary algorithms	16
Supplementary algorithm 1: MMPC algorithm over a single chromosome.	16
Supplementary algorithm 2: Identification of lead variants.	16
Supplementary methods	17
Study design and participants	17
Ethnic backgrounds	17
DNA processing and storage	17
Genotyping analysis	17
Principal component analysis and relationship inference	18
Assignment of 1KG superpopulation	18
Quality control	19

Genotype imputation	20
HLA imputation	20
Generation of genome-wide and extended chromosome-6 datasets	20
Pan-ancestry association analysis	21
Ancestry assignment	21
Principal component analysis for Europeans	21
Identification of independent associations	21
Indirect matching of independent associations	22
Meta-analysis with the study of Estrada and colleagues	22
Bayesian fine-mapping	22
Bayesian colocalisation of GWAS and eQTL signals	23
Cell type and tissue enrichment	23
Genetic risk sharing with other autoimmune diseases	23
Supplementary notes	23
Supplementary note 1	23
Supplementary note 2	24
Supplementary note 3	24
The International NMOSD Genetics Consortium	24
Description of Appendix 2 pages	27
Page 1	27
Page 2	27
Page 3	28
Page 4	28
Page 5	28
Page 6	28
Page 7	29
Page 8	29
Page 9	29
Page 10	29
Page 11	29
Page 12	29
Page 13	29
Page 14	30
Page 15	30
Page 16	30
Page 17	31

Page 18	31
Page 19	31
Page 20	32
Page 21	32
Page 22	32
Page 23	32
Page 24	32
Page 25	33
Page 26	33
Page 27	33
Reference list	33

Supplementary tables

Supplementary table 1: Centre statistics.

	Country	# samples	# cases	# controls
Guthy-Jackson Charitable Foundation	USA	863	495	368
Oxford University Hospitals Trust	UK	772	195	577
Mayo Clinic	USA	720	275	445
Accelerated Cure Project (GJCF)	USA	185	107	78
National Cancer Center	South Korea	146	146	0
Charité – Universitätsmedizin Berlin	Germany	83	35	48
Siriraj Hospital / Mahidol University	Thailand	81	79	2
Neuroscience Institute, Tehran University of Medical Sciences	Iran	78	78	0
Karolinska Institute	Sweden	60	46	14
Lyon's Neuroscience Research Center	France	52	52	0
Kyushu University	Japan	50	50	0
LMU Munich	Germany	50	34	16
University of Liverpool	UK	47	23	24
University of Düsseldorf	Germany	32	27	5
Penang General Hospital	Malaysia	31	17	14
Hospital Clinic of Barcelona	Spain	29	29	0
University of Southern Denmark	Denmark	28	28	0
Medical University of Innsbruck	Austria	27	6	21
Kyoto Min-iren Central Hospital	Japan	27	27	0
University of Buenos Aires	Argentina	27	17	10
University of Pécs / University of Southern Denmark	Hungary	25	25	0
Medical University of Vienna	Austria	20	11	9
Ruhr University Bochum	Germany	20	14	6
Münster University Hospital	Germany	19	9	10
Neurology Department, Technical University of Munich	Germany	18	16	2
Neurology Clinic, Clinical Centre of Serbia	Serbia	17	17	0
Vall d'Hebron Hospital (Cemcat, the Multiple Sclerosis Center of Catalonia)	Spain	16	15	1
Faculty of Health Sciences, University Fernando Pessoa	Portugal	13	6	7
University of Cardiff	UK	13	9	4
University Hospital Heidelberg	Germany	10	6	4
Institute of Neurology, Fondazione Policlinico A. Gemelli IRCCS, Catholic University of Sacred Heart	Italy	7	7	0
University Hospital Basel	Switzerland	7	0	7
University of Helsinki	Finland	6	5	1
Stellenbosch University	South Africa	5	5	0
University of Porto	Portugal	5	5	0
KU Leuven	Belgium	3	3	0

Centres are sorted by the number of samples provided.

Supplementary table 2: Disease onset pattern statistics.

Optic neuritis	Transverse myelitis	Brainstem / cerebral	Optic neuritis + transverse myelitis	Other mixed	Unknown
315 (20.0%)	352 (22.4%)	48 (3.1%)	238 (15.1%)	605 (38.5%)	15 (1.0%)

Supplementary table 3: Initial attack recovery statistics.

Poor	Partial	Complete	Unknown
141 (9.0%)	302 (19.2%)	266 (16.9%)	864 (54.9%)

Supplementary table 4: Ethnic background statistics.

	# samples	# 1KG outliers
White	2139	40
Asian	480	22
Unknown	297	14
Black or African American	248	14
Hispanic or Latino	183	13
Greater Middle Eastern	85	78
Mixed	75	21
American Indian or Alaska native	12	1
Other	12	1
Native Hawaiian/Pacific Islander	2	0

Ethnic backgrounds are sorted by number of samples.

Supplementary table 5: Number of samples per ethnic background and 1KG superpopulation.

	European	East Asian	African	American	South Asian
White	2089	0	6	32	12
Asian	30	395	2	2	51
Unknown	233	12	34	12	6
Black or African American	2	2	234	10	0
Hispanic or Latino	57	1	5	120	0
Greater Middle Eastern	76	0	0	1	8
Mixed	16	2	14	37	6
American Indian or Alaska native	3	4	1	1	3
Other	0	5	4	1	2
Native Hawaiian/Pacific Islander	0	2	0	0	0

Rows are ethnic backgrounds, sorted by total number of samples. Columns are 1KG superpopulations, sorted by total number of samples.

Supplementary table 6: Association statistics for each study.

	# cases	# controls	# variants	λ_{gc}	# independent associations ($p \leq 10^{-5}$)	# independent associations ($p \leq 5 \times 10^{-8}$)
PAN	1573	1260	6264966	0.999	17	3
EUR	803	1054	6007187	0.995	16	3

Supplementary table 7: Matching coefficients of selected autoimmune diseases with AQP4-positive NMOSD.

	Short name	Accession	# independent associations	Matches with AQP4-positive NMOSD	Matching coefficient with AQP4-positive NMOSD
Myositis	Myositis	GCST003092	9	rs3130614 ↔ rs1270942 rs2734583 ↔ rs1270942 rs9267488 ↔ rs1270942 rs3131619 ↔ rs1270942 rs3094013 ↔ rs1270942 rs3099844 ↔ rs1270942 rs3094011 ↔ rs1270942	0.32
Systemic lupus erythematosus	SLE	GCST003103	7	rs1150753 ↔ rs1270942 rs3131379 ↔ rs1270942 rs9267531 ↔ rs1270942	0.17
Idiopathic membranous nephropathy	IMN	GCST000984	21	rs7775397 ↔ rs1270942 rs389884 ↔ rs1270942 rs3134792 ↔ rs1270942 rs7750641 ↔ rs1270942 rs3130544 ↔ rs1270942	0.16
Collagenous colitis	CC	GCST90204145	1	rs2844531 ↔ rs1270942	0.12
Sarcoidosis (Lofgren's syndrome)	Sarcoidosis	GCST005540	1	rs3130288 ↔ rs1270942	0.12
Microscopic colitis	MC	GCST90204144	2	rs9267445 ↔ rs1270942	0.11
Psoriasis	Psoriasis	GCST000165	2	rs3134792 ↔ rs1270942	0.11
Systemic sclerosis	SSc	GCST000650	4	rs3821236 ↔ rs3821236	0.1
Myasthenia gravis	MG	GCST001611	5	rs3130544 ↔ rs1270942	0.1
Common variable immunodeficiency	CVID	GCST90301675	6	rs3117577 ↔ rs1270942	0.09
Type 1 diabetes and autoimmune thyroid diseases	T1D&AITD	GCST002876	11	rs1270942 ↔ rs1270942	0.07
Primary sclerosing cholangitis	PSC	GCST006670	13	rs4143332 ↔ rs1270942	0.07
Sjögren's syndrome	SjS	GCST004878	15	rs3135394 ↔ rs1270942	0.06
Rheumatoid arthritis	RA	GCST005569	50	rs13426947 ↔ rs3821236	0.03
Multiple sclerosis	MS	GCST005531	122		0

Statistics are based on the GWAS-Catalog studies in Europeans with the greatest matching coefficient with our AQP4-positive NMOSD study in Europeans. Diseases are sorted by matching coefficient with AQP4-positive NMOSD.

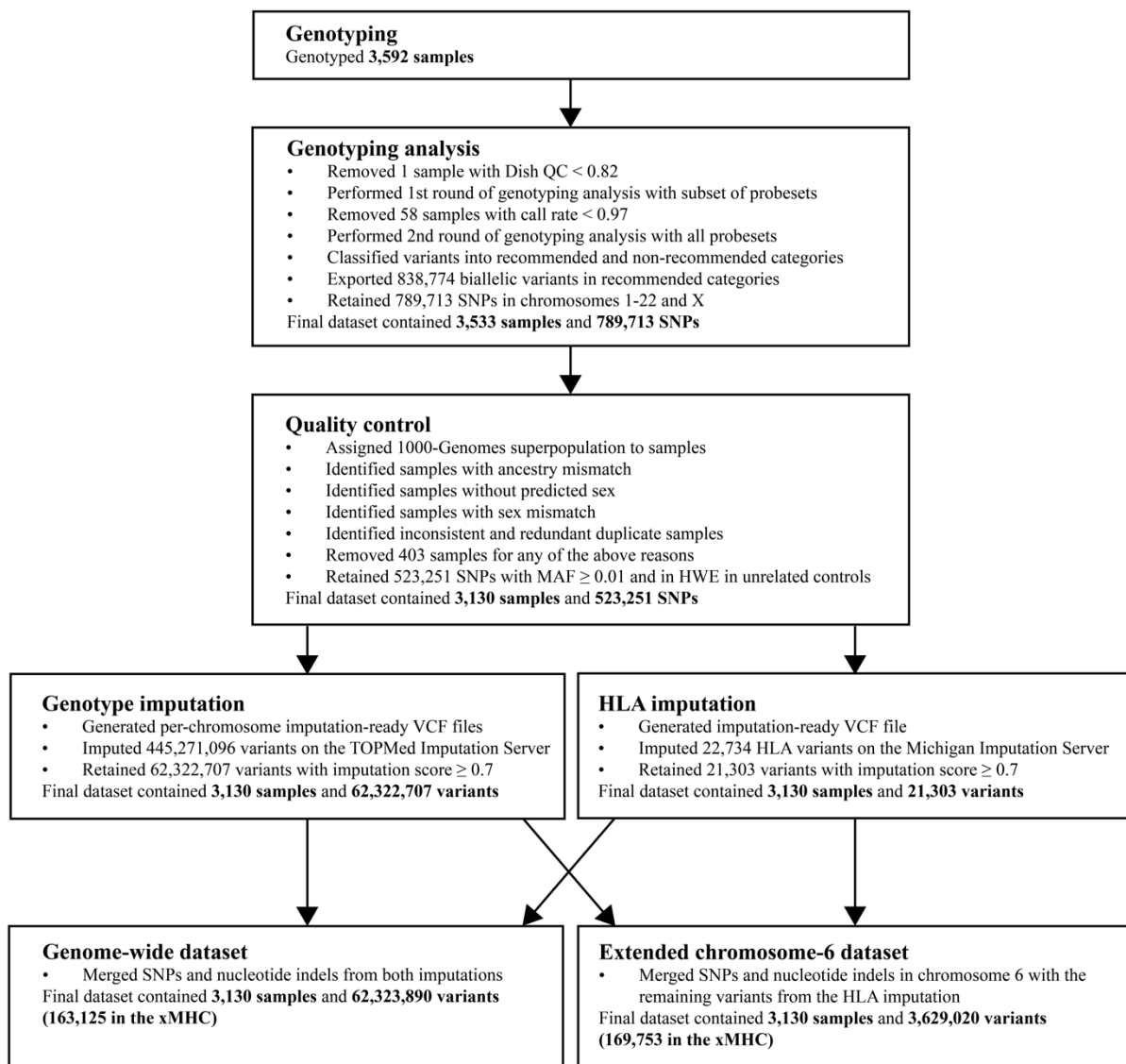
Supplementary table 8: Correlations of selected autoimmune diseases with AQP4-positive NMOSD.

	Short name	Accession	Correlation	95% CI	P-value
Sarcoidosis (Lofgren's syndrome)	Sarcoidosis	GCST005540	0.81	[0.54, 1.00]	6.081 x 10 ⁻⁹
Systemic lupus erythematosus	SLE	GCST003156	0.77	[0.45, 1.00]	1.957 x 10 ⁻⁶
Rheumatoid arthritis	RA	GCST005569	0.11	[0.00, 0.37]	0.3995
Collagenous colitis	CC	GCST90204145	0.49	[0.00, 1.00]	0.4604
Microscopic colitis	MC	GCST90204144	0.39	[0.00, 1.00]	0.6794
Multiple sclerosis	MS	GCST005531	< 0.01	[0.00, 0.19]	0.9898

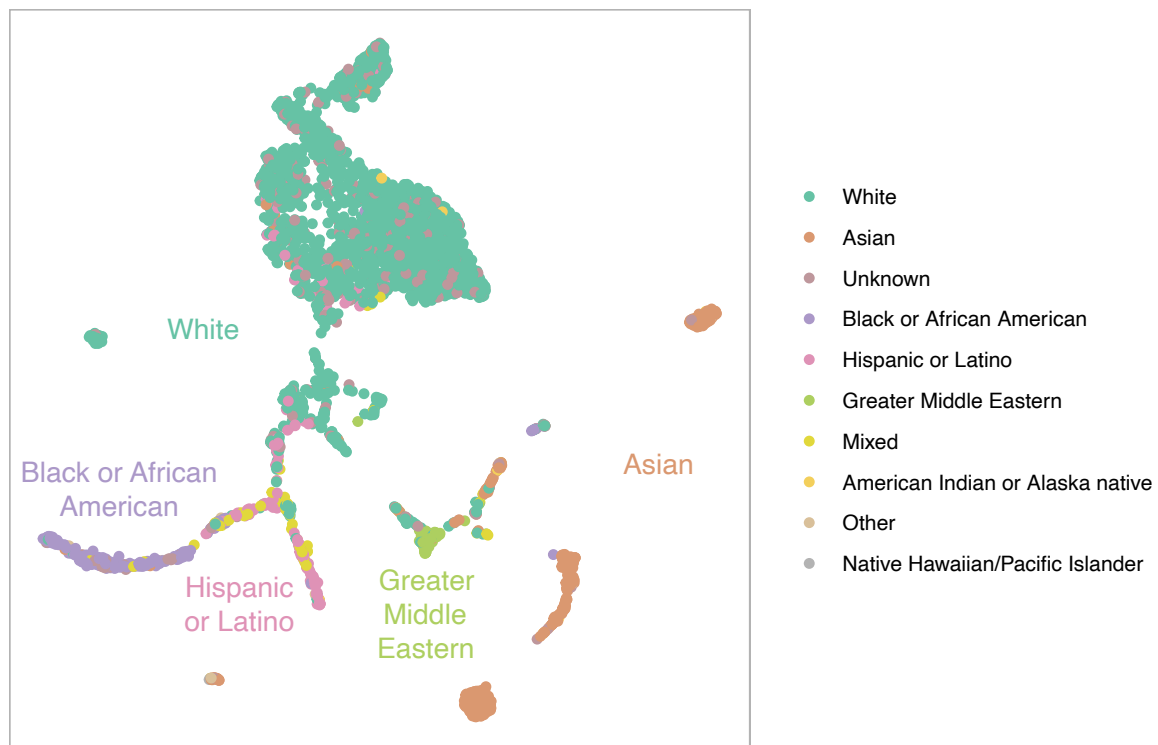
Diseases are sorted by p-value.

Supplementary figures

Supplementary figure 1: Flow chart of the process of generating the genome-wide and extended chromosome-6 datasets.

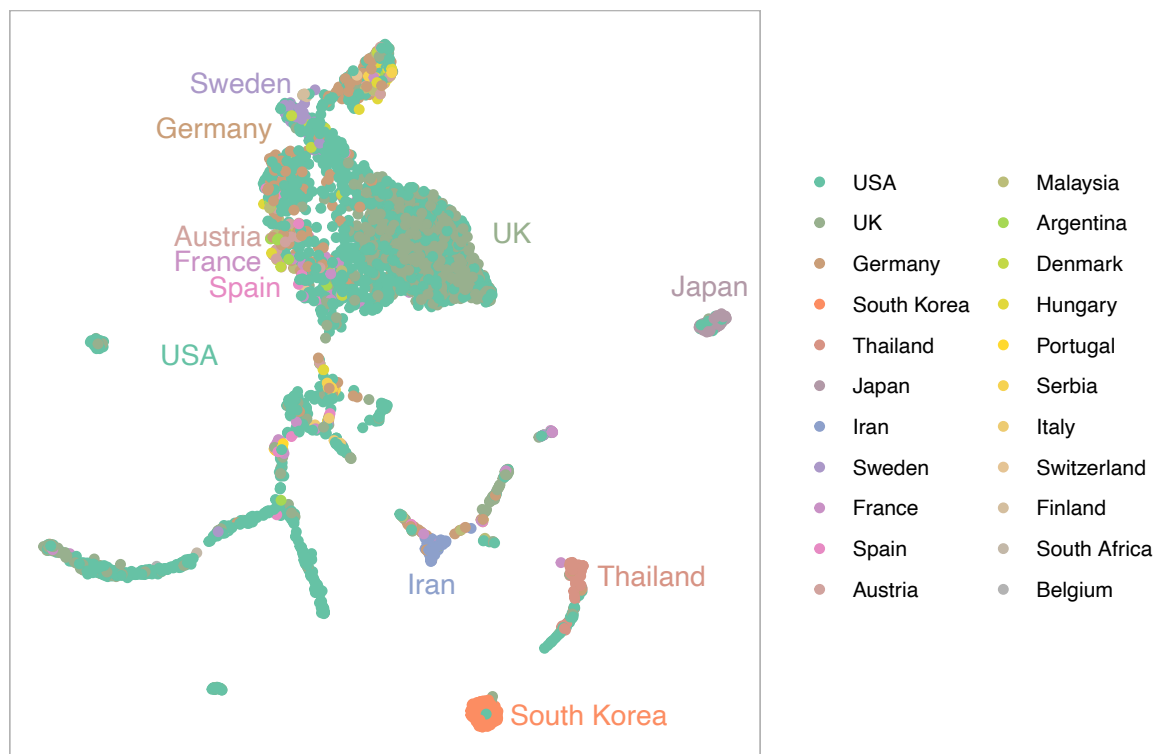


Supplementary figure 2: UMAP plot of the samples, coloured by ethnic background.



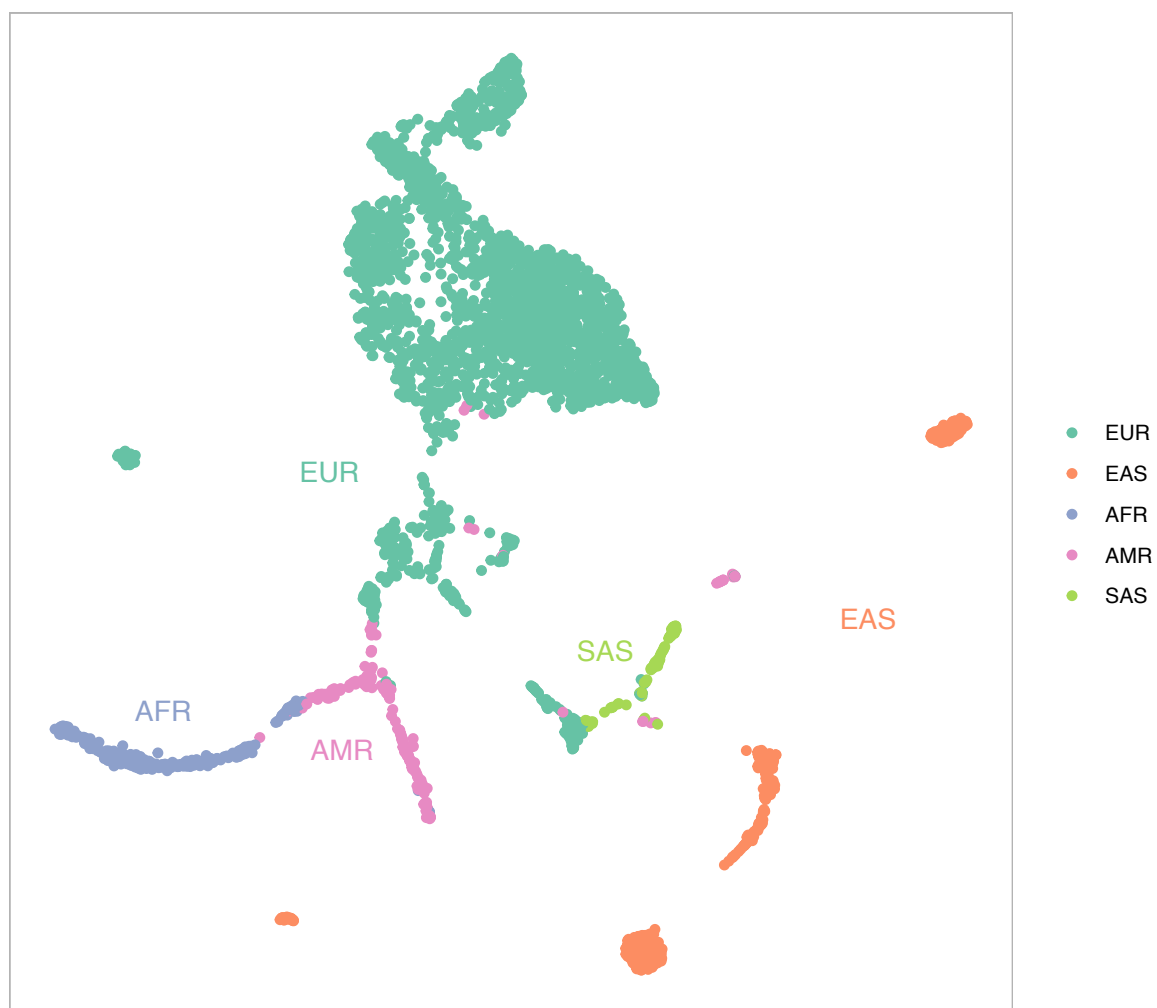
Ethnic backgrounds are ordered by number of samples. Top ethnic backgrounds other than Mixed, Other, and Unknown are labelled in the plot.

Supplementary figure 3: UMAP plot of the samples, coloured by country of centre of origin.



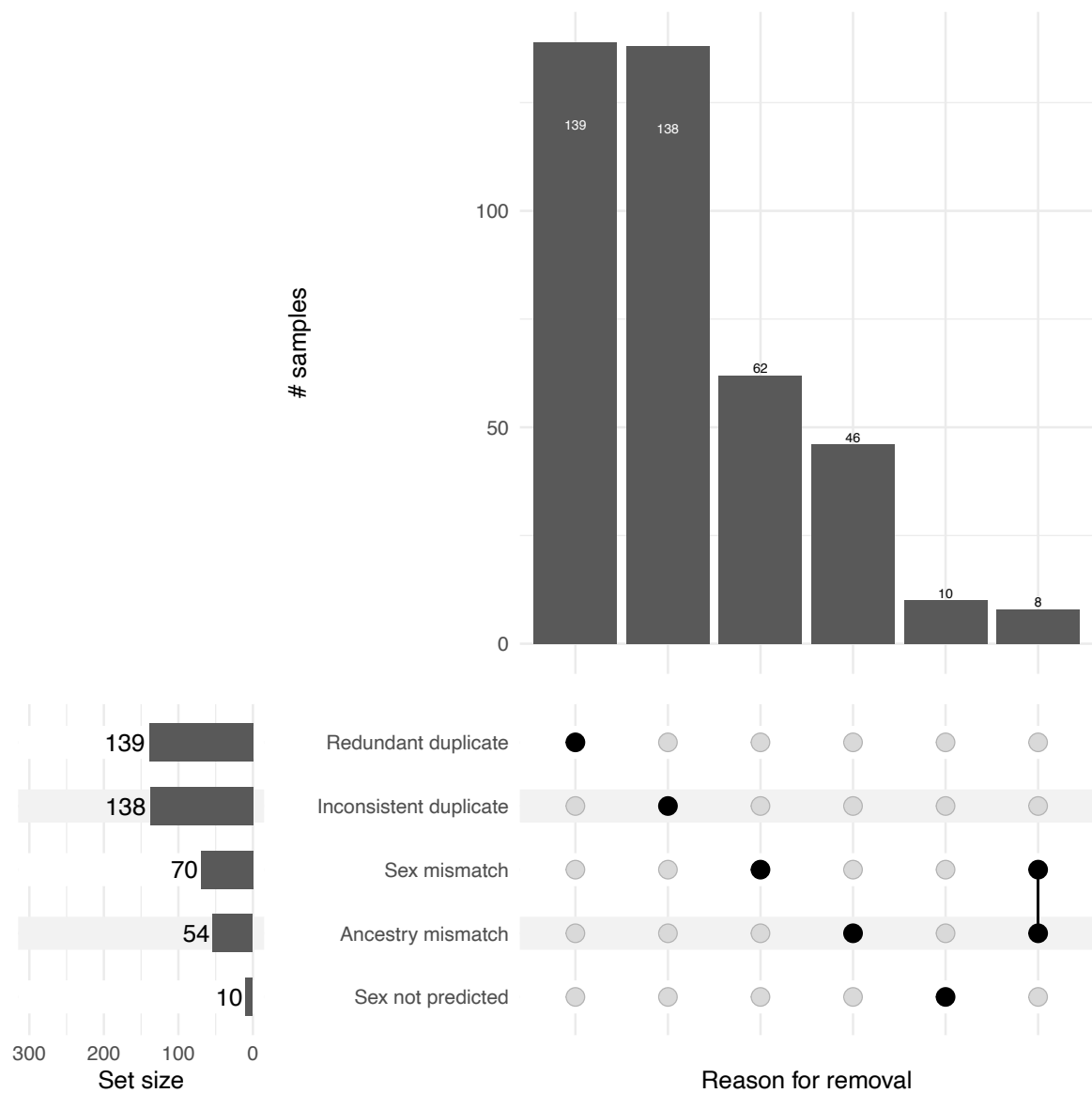
Countries are ordered by number of samples. The top 50% countries are labelled in the plot.

Supplementary figure 4: UMAP plot of the samples, coloured by 1KG superpopulation.

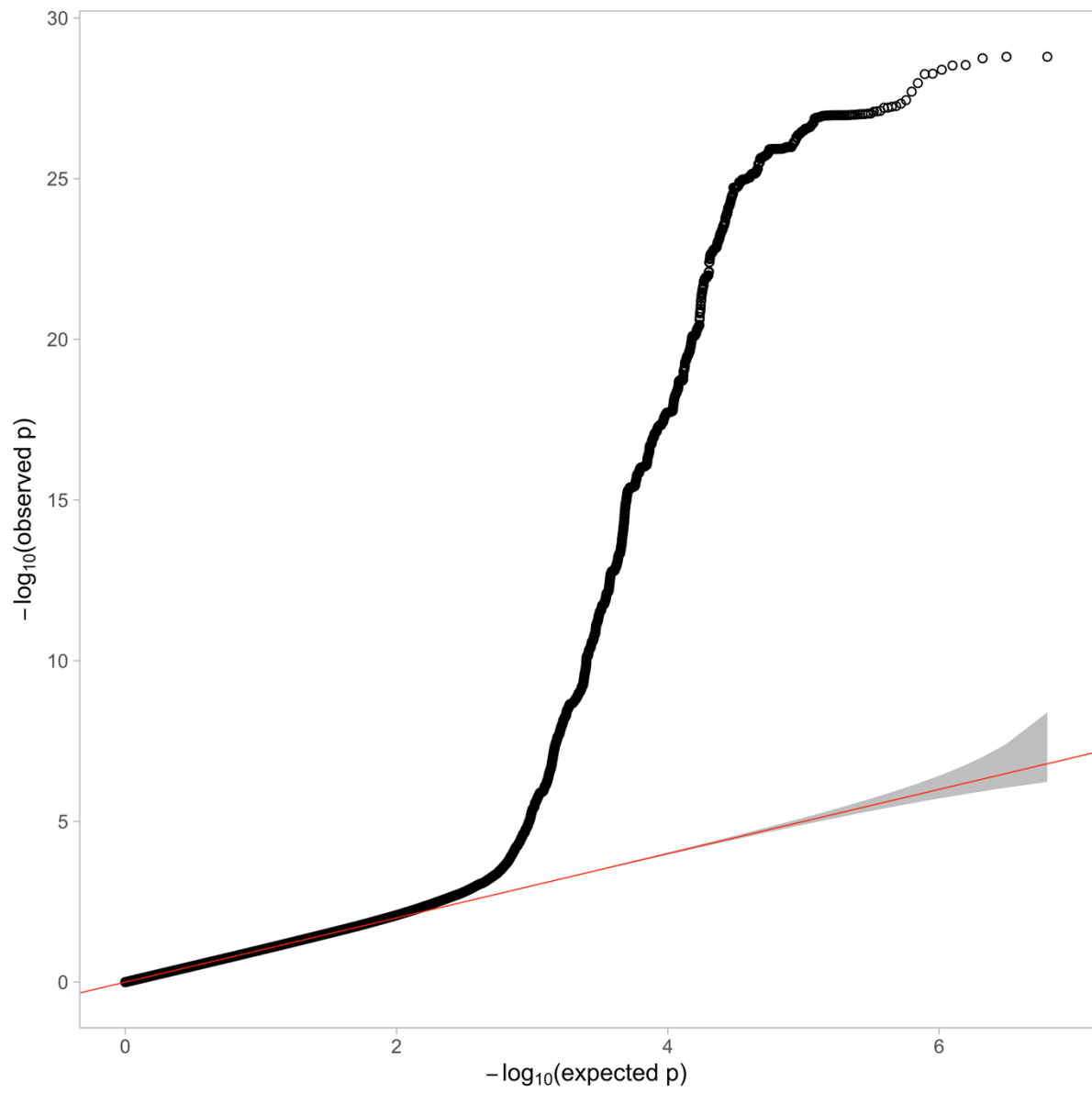


1KG superpopulations are ordered by number of samples.

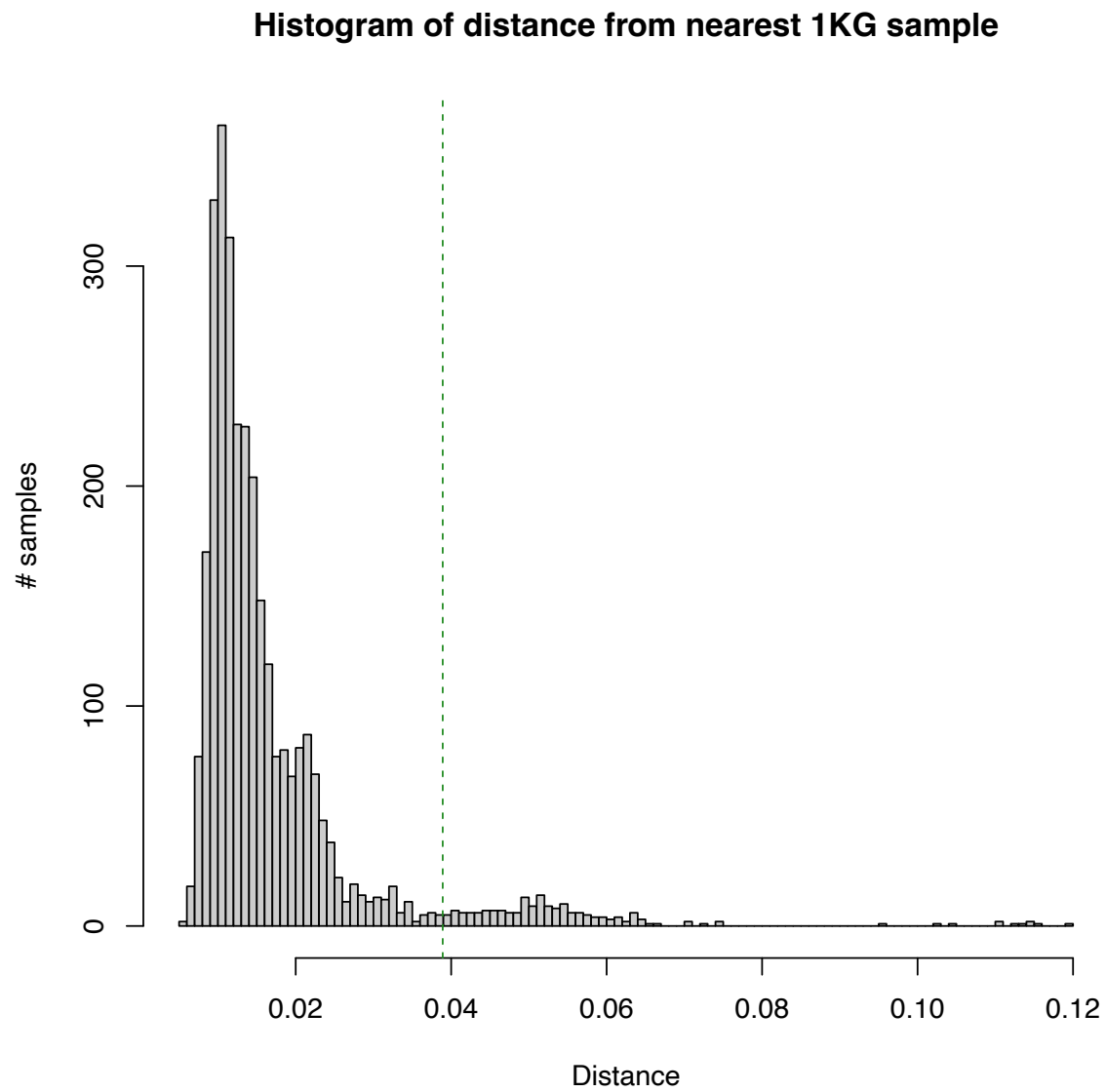
Supplementary figure 5: UpSet plot of sample reasons for removal.



Supplementary figure 6: QQ plot for the pan-ancestry study.

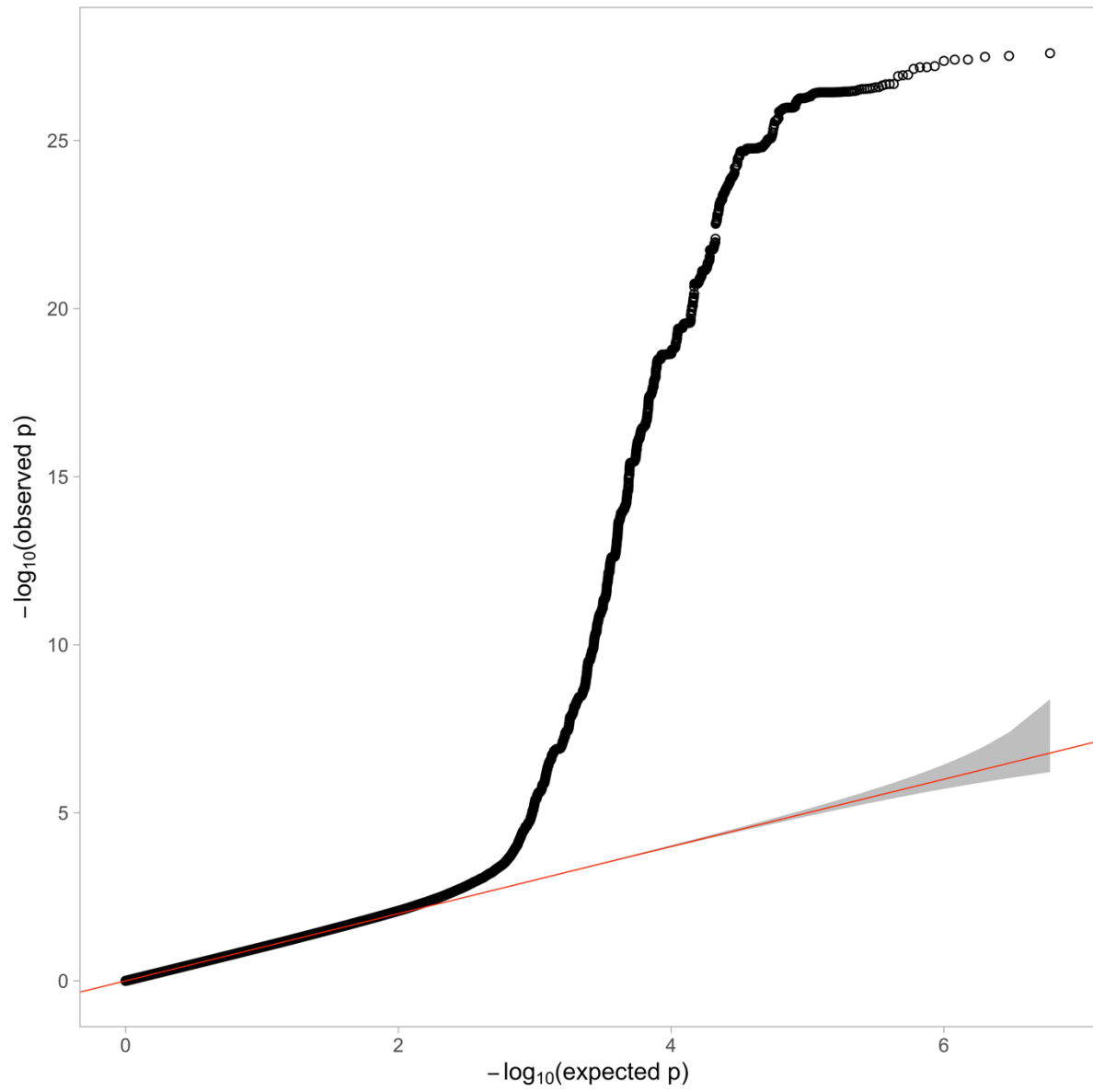


Supplementary figure 7: Histogram of sample distance from the nearest 1KG sample.

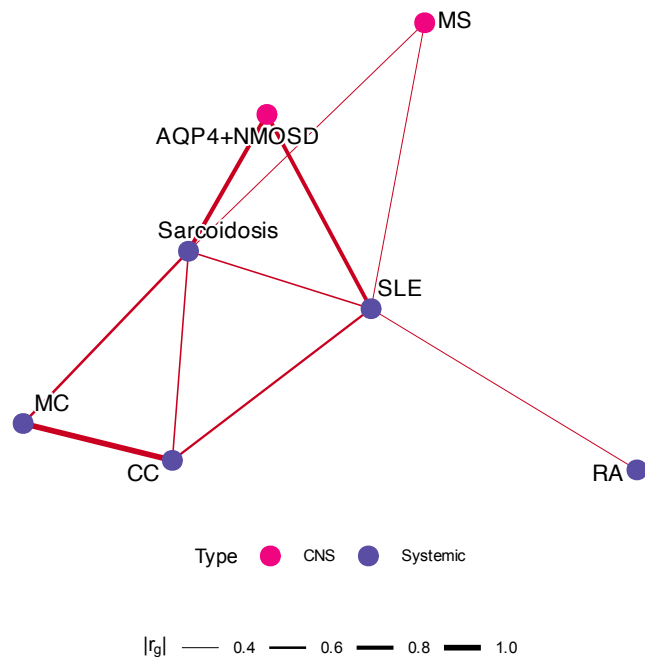


The vertical dashed green line shows the threshold beyond which samples were considered outliers.

Supplementary figure 8: QQ plot for the European study.

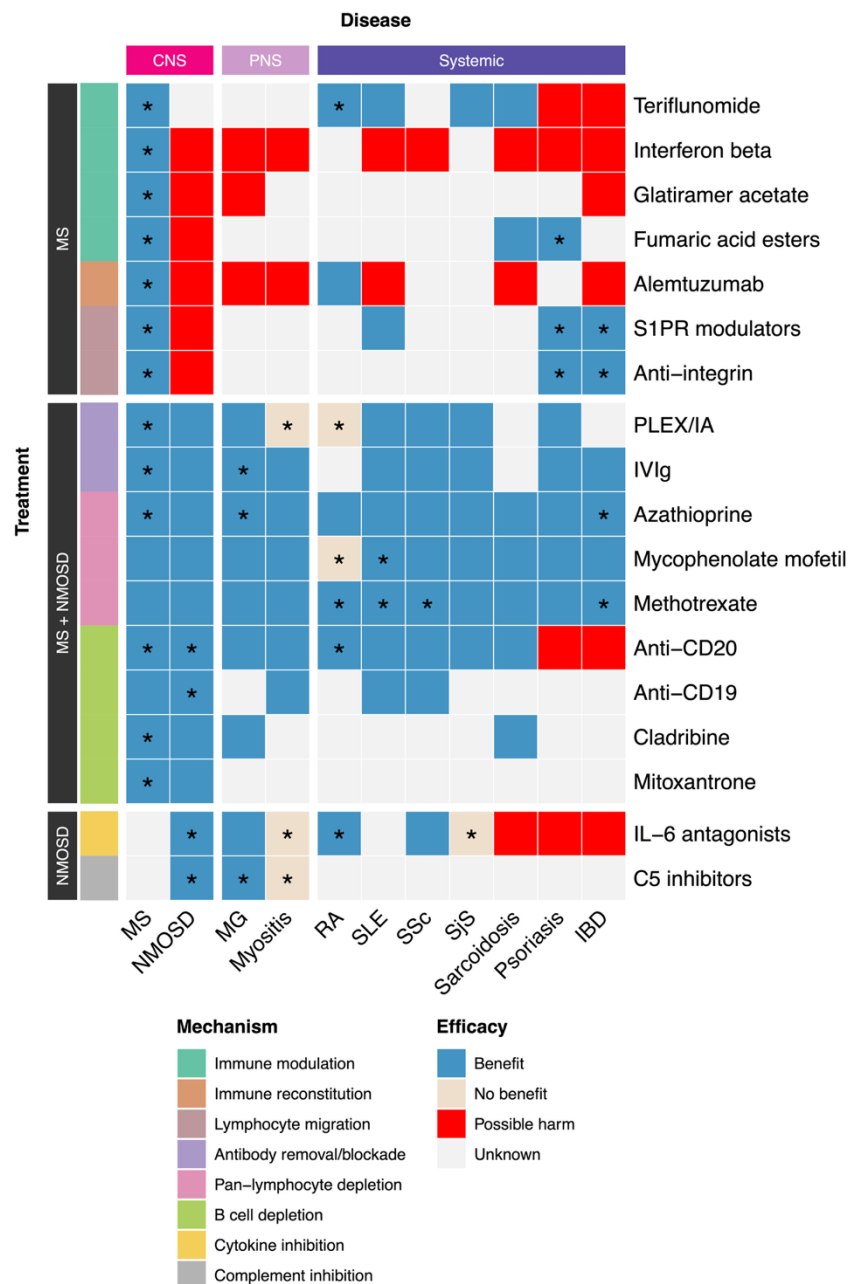


Supplementary figure 9: Graph of genetic correlations.



Nodes comprise AQP4-positive MOSD, autoimmune diseases with matches to AQP4-positive MOSD and summary statistics available, and multiple sclerosis (See supplementary table 8). Nodes with nominally significant correlation are connected by an edge whose thickness is proportional to the correlation, while the distance between them is approximately inversely proportional to the correlation. Nodes without edges were removed. AQP4+NMOSD = AQP4-positive NMOSD, CC = collagenous colitis, MC = microscopic colitis, MS = multiple sclerosis, RA = rheumatoid arthritis, SLE = systemic lupus erythematosus.

Supplementary figure 10: Heatmap of comparative treatment efficacy across selected CNS, PNS, and systemic autoimmune diseases.



Current non-steroid treatments in clinical use for the acute or chronic management of multiple sclerosis and/or NMOSD, grouped by mechanism of action and evidence of efficacy in either multiple sclerosis, NMOSD, or both. Treatment efficacy is categorised as “benefit”, “no benefit”, “possible harm” or “unknown” based on a review of published trials and/or case series for each of multiple sclerosis, NMOSD, and a selection of peripheral nervous system (PNS) and systemic autoimmune diseases which are known comorbidities of NMOSD. Asterisks (*) indicate evidence of efficacy from randomised controlled trials and/or meta-analyses. IBD = inflammatory bowel disease, MG = myasthenia gravis, MS = multiple sclerosis, RA = rheumatoid arthritis, SjS = Sjögren's syndrome, SLE = systemic lupus erythematosus, SSc = systemic sclerosis.

Supplementary algorithms

Supplementary algorithm 1: MMPC algorithm over a single chromosome.

MMPC algorithm over a single chromosome

Input: Set of variants V in a single chromosome, phenotype P , covariates C

Output: Set of variants I independently associated with P

```
1 initialise  $I$  with the empty set
2 initialise set OPEN (variants considered for insertion in  $I$ ) with  $V$ 
3 // forward selection phase
4 repeat (until OPEN is empty)
5     test the conditional independence (CI) of each variant  $X$  in OPEN with  $P$  given  $C$  and each
       subset  $Z$  of  $I$  such that (a) all variants in  $Z$  are in the xMHC or within 1 mb of  $X$ , if  $X$  is in the
       xMHC, or (b) all variants in  $Z$  are within 1 mb of  $X$ , if  $X$  is not the xMHC
6     remove variants from OPEN that become independent from  $P$  given  $C$  and some subset of  $I$ 
7     select variant  $X$  in OPEN with the minimum maximum p-value across all CI tests that were
       performed for  $X$ 
8     move  $X$  from OPEN to  $I$ 
9 end
10 // backward elimination phase
11 for (each variant  $X$  in  $I$ )
12     remove  $X$  from  $I$  if there is a subset  $Z$  of the other variants in  $I$  such that (a) all variants in  $Z$ 
       are in the xMHC or within 1 mb of  $X$ , if  $X$  is in the xMHC, or (b) all variants in  $Z$  are within 1
       mb of  $X$ , if  $X$  is not the xMHC, and  $X$  is independent from  $P$  given  $C$  and  $Z$ 
13 end
```

Supplementary algorithm 2: Identification of lead variants.

Identification of lead variants

Input: Set of associated variants V

Output: Set of lead variants I

```
1 initialise  $I$  with the empty set
2 initialise set OPEN (variants considered for insertion in  $I$ ) with  $V$ 
3 repeat (until OPEN is empty)
4     identify variant  $X$  with the minimum p-value in OPEN
5     remove variants from OPEN that are (a) 1 mb around  $X$ , if  $X$  is not in the xMHC, or (b) 1 mb
       around  $X$  or in the xMHC, if  $X$  is in the xMHC
6     move  $X$  from OPEN to  $I$ 
7 end
```

Supplementary methods

Study design and participants

From 2013, DNA samples from patients with NMOSD or controls (comprising non-affected relatives, healthy non-relatives and those with other autoimmune diseases) were curated from across the globe, with the last sample received in February 2022, amassing 3,592 samples from 36 contributing centres (supplementary table 1). International DNA samples from de-identified AQP4-positive NMOSD and control donors were obtained from each contributing hospital or research facility under their local approvals. UK samples were obtained under two NHS REC authorisations held by Oxford University Hospital trust (NHS REC 14/SC/0057 and NHS REC 21/SC/0353). All NMOSD patients had to fulfil the 2015 NMOSD criteria for inclusion,¹ including AQP4-IgG status. Additional information collected for each sample included age at onset, sex, comorbid disease (yes or no), ethnic background, disease presentation (optic neuritis, transverse myelitis, and/or brainstem/cerebral involvement), as well as recovery (poor, partial, or complete) (table, supplementary tables 2 and 3). Samples of patients with CNS autoimmune diseases were processed along with the rest of the samples but were not used as controls in any of our analyses.

Ethnic backgrounds

The ethnic background of the samples was recorded as Hispanic or Latino, American Indian or Alaska native, Black or African American, Asian, White, Mixed, Other – specify, or Unknown. Two new ethnic backgrounds were defined for the analysis, Greater Middle Eastern and Native Hawaiian/Pacific Islander. Samples from Iran (all of which were reported as White) and samples of “Other” background specified as Egyptian, Iranian, Jewish, Lebanese, Moroccan, or Syrian were reclassified as Greater Middle Eastern,² while samples of “Other” background specified as Native Hawaiian/Pacific Islander were reclassified as such.

DNA processing and storage

DNA was either processed at source and sent to Oxford for quantification or isolated from blood samples in Oxford using the QIAamp Blood DNA Kit. DNA was quantified in duplicate using both the Quant-iT™ dsDNA Assay Kit and the NanoDrop spectrophotometer. A standard curve was included on each plate to ensure consistency across measurements. Additionally, samples from each centre were run on a 1% agarose gel to assess DNA integrity and confirm that the fragments were greater than 10 kb in size. All DNA samples were diluted using reduced EDTA TE buffer (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA), supplied by Affymetrix, to achieve a final DNA concentration of 5 ng/μl. DNA was stored at -80 C before being shipped on dry ice for genotyping.

Genotyping analysis

Genotyping was performed at Cambridge Genomic Services at University of Cambridge using the Axiom Human Genotyping SARS-Cov-2 Research Array, which was selected for increased coverage of genomic regions that regulate immune function, while still achieving genome-wide coverage. The Best Practices Genotyping Analysis Workflow from the Axiom Genotyping Solution Data Analysis User Guide (Publication Number MAN0018363, Revision B.0) was applied using Analysis Power Tools (APT; v2.11.6) to perform genotyping of the samples in a single batch (supplementary figure 1). Specifically, one sample with Dish QC < 0.82 was first removed. A first round of genotyping was performed with the remaining 3,591 samples for a collection of previously wet-lab-tested probe sets; 58 samples with call rate < 0.97 were removed. The mean passing-sample call rate was ≥ 0.985 for all plates, so no plate was removed. A second round of genotyping was performed with the remaining 3,533 samples for all probesets. Variants were then classified into a set of recommended and non-recommended categories based on the best performing probesets (note that inclusion in the former set implies a call rate of ≥ 0.95). The cluster plots of 200 randomly selected variants from each category were visually inspected. Finally, the genotypes of the 838,774 biallelic variants in recommended categories were exported to the PLINK TPED format and the 789,713 SNPs in chromosomes 1-22 and X were retained.

To be used in R, the dataset was converted to the GDS format by the `snpGDSBED2GDS` function in the `SNPRelate` R package³ (v1.34.1) with option `snpGDSOption(25 = 24)`, as the X-chromosome pseudo-autosomal region (PAR) is represented by 25 in PLINK 1.9 but 24 in GDS.

Principal component analysis and relationship inference

Initial pairwise sample kinship coefficients, subsequently referred to as KING, were computed using the KING-robust estimator⁴ (`--make-king`) in PLINK 2⁵ (v2.00a5.12). Increasingly-better principal components and kinship coefficients were then computed for the samples using the PC-AiR⁶ and PC-Relate⁷ methods in succession for two rounds.

In the first round, KING were supplied both as kinship coefficients and ancestry divergence measures to the `pcairPartition` function in the GENESIS R package⁸ (v2.30.0, with the fix in <https://github.com/UW-GAC/GENESIS/issues/77>) to partition the dataset into a set of samples that are mutually unrelated and representative of the ancestries in the dataset (“unrelated” samples, subsequently referred to as UNRELATED_1) and a set of samples that are related to samples in the first set (“related” samples).⁶ SNPs with minor-allele frequency (MAF) ≥ 0.05 among UNRELATED_1 were identified by the `--maf 0.05` command in PLINK 1.9⁹ (v1.90b6.21), and Hardy-Weinberg equilibrium (HWE) statistics for those SNPs were calculated from UNRELATED_1 by the `--hardy` command. Inbreeding coefficients were then calculated from the output of that command as $1 - O(HET_A1)/E(HET_A1)$, and SNPs with inbreeding coefficient < -0.2 (corresponding to an excess of heterozygotes) were removed; SNPs with inbreeding coefficient > 0.2 (corresponding to heterozygote depletion) were not removed, as they are expected from population stratification,⁹ which had been not yet dealt with at that point. Inbreeding coefficients were used instead of p-values, as the latter depend on the sample size.¹⁰ Using UNRELATED_1, the remaining SNPs were LD-pruned by `--indep-pairwise 5000 kb 0.1`, and a principal component (PC) transformation over the LD-pruned SNPs with 20 PCs was learned by `--pca` in PLINK 2. The LD-pruning parameters were chosen so that no peaks would be observed in the Manhattan plot of the SNP loadings of each PC.¹¹ The association of each PC X with each (a) ethnic background other than Mixed, Other, and Unknown, (b) centre, and (c) country Y with ≥ 20 samples was then tested as follows. First, X was standardised. Let μ be the difference in mean X between samples in Y and samples not in Y , and $\hat{\mu}$ be the estimate of μ . The p-value of the hypothesis that $|\mu| \geq 1$ was obtained by performing a t-test of the hypothesis that $\mu \geq 1$, if $\hat{\mu} \geq 0$, and $\mu \leq -1$ otherwise, and doubling its p-value and capping it at 1. The number of PCs retained was then set to the order of the highest order PC for which a significant association ($p \leq 10^{-3}$) between the PC and a covariate was discovered, which was PC12, associated with Sweden. The PC transformation with 12 PCs was applied to the entire dataset by `--score` in PLINK 2, as described in <https://www.cog-genomics.org/plink/2.0/score> (“PCA projection with `--score`”), with its results subsequently referred to as PCS_1. Ancestry-corrected kinship coefficients (subsequently referred to as PC_RELATE_1) were computed by `GENESIS::pcrelate()` using the LD-pruned SNPs, PCS_1, and UNRELATED_1 as the training set, followed by `GENESIS::pcrelateToMatrix()` with `scaleKin = 1`.⁷

In the second round, a set of unrelated samples (UNRELATED_2) was identified by `GENESIS::pcairPartition()` given PC_RELATE_1 and using KING as ancestry divergence measures. SNPs with MAF ≥ 0.05 among UNRELATED_2 were identified, and ancestry-corrected inbreeding coefficients were computed for them by RUTH¹² with parameters `--lambda 0 --lrt-em`, given PCS_1. SNPs with inbreeding coefficient < -0.2 or > 0.2 were removed. The remaining SNPs were LD-pruned and a PC transformation over the LD-pruned SNPs with 20 PCs was learned. PC12 was again found to be the highest-order PC significantly associated with a covariate, so the PC transformation with 12 PCs was applied to the entire dataset to generate PCS_2. Finally, kinship coefficients (PC_RELATE_2) were computed by `GENESIS::pcrelate()` using the LD-pruned SNPs, PCS_2, and UNRELATED_2 as the training set, followed by `GENESIS::pcrelateToMatrix()` with `scaleKin = 1`.

In the end, a set of unrelated samples (UNRELATED_3) was identified using PC_RELATE_2 (and KING as ancestry divergence measures) and a UMAP embedding of the samples was computed from PCS_2 by the `umap` function in the R `uwot` package (v0.1.16) with parameters `n_neighbors = 50` and `spread = 5` (supplementary figures 2-4).

Assignment of 1KG superpopulation

Ethnic background was missing for 297 samples (supplementary table 4), prohibiting their inclusion in ancestry-specific analyses. To maximise the number of samples used in those analyses and validate the reported ethnic backgrounds, a 1000 Genomes¹³ (1KG) superpopulation was assigned to the samples using the 1KG dataset.

The samples from phase 3 of The 1000 Genomes Project were first downloaded in PLINK format from https://www.cog-genomics.org/plink/2.0/resources#phase3_1kg. The dataset was filtered to remove samples belonging to more than one 1KG population and variants that are not autosomal SNPs. SNPs were renamed by `--set-all-var-ids @:#:$1:$2` in PLINK 2. Two rounds of PC-AiR and PC-Relate were performed as described

above, with the results of the second round referred to as 1KG_PCS_2 and 1KG_PC_RELATE_2. The 1KG populations were the covariates used to decide the number of PCs in each round (22 in round 1, PC22 being significantly associated with ITU, and 20 in round 2, PC20 being significantly associated with BEB).

For each population, a set of unrelated samples was identified given 1KG_PC_RELATE_2 and used to calculate the population's median-based centroid in the second PC space. The distance of each sample from the centroid was then calculated and the median absolute deviation (MAD) from the median distance was calculated using the unrelated samples. Samples with distance more than 3 MADs away from the median were considered outliers and were discarded. A set of unrelated samples (1KG_UNRELATED_3) was identified among the remaining samples in the dataset.

After renaming the SNPs in our dataset using `--set-all-var-ids @:#:\$1:\$2` in PLINK 2, the SNPs shared between our dataset and the 1KG dataset were identified. A merged dataset was then created by PLINK `--bmerge` over the SNPs present in both datasets. SNPs with $MAF \geq 0.05$ among 1KG_UNRELATED_3 were identified. Among those SNPs, SNPs in HWE in the 1KG dataset were identified using RUTH with 1KG_UNRELATED_3 and 1KG_PCS_2, and SNPs in HWE in our dataset were identified using RUTH with UNRELATED_3 and PCS_2. SNPs in HWE in both datasets were LD-pruned. A PC transformation over the LD-pruned SNPs with 20 PCs was learned from 1KG_UNRELATED_3 and applied to all samples. Finally, every sample in our dataset was assigned the 1KG superpopulation of the nearest 1KG sample in the PC space.

Quality control

First, samples whose sex was not predicted by APT and mislabelled samples were identified. Mislabelled samples comprised samples with inconsistent sex (sex was both reported and predicted, but reported and predicted sex differed) or inconsistent ancestry (ethnic background was reported and was inconsistent with the assigned 1KG superpopulation). Ancestry was deemed consistent in the following cases:

- Ethnic background was Hispanic or Latino, as individuals identifying as such are genetically diverse¹⁴ and 57 (31.1%) of those samples were indeed assigned the European 1KG superpopulation. We note that all 16 samples from Cemcat were reported as Hispanic or Latino, with 13 (81.3%) of them assigned European. The ethnic background of the 29 HCB samples was not available, but 26 (89.7%) of them were assigned European. Finally, all 27 samples from Argentina were also reported as Hispanic or Latino and 24 (88.9%) of them were assigned European.
- Ethnic background was American Indian or Alaska native, Mixed, Other, Greater Middle Eastern, or Native Hawaii/Pacific Islander, neither of which is represented in 1KG.
- 1KG superpopulation was American, which is an admixed population and whose assigned samples are from various ethnic backgrounds, as seen in supplementary table 5.
- Ethnic background was Black or African American and 1KG superpopulation was African.
- Ethnic background was Asian and 1KG superpopulation was East Asian or South Asian.
- Ethnic background was White and 1KG superpopulation was European.

Sample pairs with final kinship coefficient ≥ 0.354 were deemed to be duplicates. Groups of duplicate samples were identified by finding connected components in the undirected graph with nodes being the samples in duplicate-sample pairs and edges connecting such pairs. Each group of duplicate samples was processed as follows; at the end of each step, if there were < 2 samples remaining in the group, processing of the group stopped:

- Mislabelled samples were removed from the group.
- Remaining samples were checked for consistency. For the samples to be considered consistent, in the sense that there is no evidence that they came from different individuals, they had to be in the same category. Ethnic background, year of birth, month of birth, and day of birth had to match when present in more than one samples. If age at onset was reported for some samples, it had to be the same; if there were also samples for which age at onset was estimated, estimated values had to be within ± 1 year of the reported value. If there were only samples with estimated values, the values had to match exactly. If the samples were inconsistent, they were marked as inconsistent duplicates, and the group was emptied.
- Remaining samples other than the one with the highest call rate were marked as redundant duplicates.

Note that the reason that mislabelled samples were not removed before duplicate detection was to obtain a report of all duplicates for our records.

Finally, samples whose sex was not predicted, mislabelled samples, inconsistent duplicates, and redundant duplicates (403 samples in total) were removed from the study, leaving 3,130 samples (supplementary figure 5).

A set of unrelated samples (UNRELATED_FILTERED) was identified among the remaining samples, and SNPs with $MAF \geq 0.05$ in that set were identified. Among those SNPs, SNPs in HWE were identified using RUTH with UNRELATED_3 and PCS_2 (in order to use as much information as possible) and LD-pruned using UNRELATED_FILTERED. A PC transformation over the LD-pruned SNPs with 20 PCs was learned from UNRELATED_FILTERED. The number of PCs to retain was once again determined to be 12, so the PC transformation with 12 PCs was applied to all remaining samples to generate PCS_FILTERED.

A set of unrelated samples (UNRELATED_FILTERED_CONTROL) among the remaining control samples was identified, and SNPs with $MAF \geq 0.01$ in that set were found. Among those SNPs, SNPs in HWE were identified using RUTH with UNRELATED_FILTERED_CONTROL and PCS_FILTERED. Those 523,251 SNPs were retained in the final pre-imputation dataset.

Genotype imputation

A PLINK dataset over the remaining SNPs and samples was first created by --make-bed in PLINK 1.9, using --merge-x to merge the non-PAR and PAR of chromosome X. SNP MAFs were then calculated by --freq in PLINK 1.9. Per-chromosome imputation-ready VCF files were subsequently generated by the script produced by the HRC-1000G-check-bim.pl script (v4.3.0) from McCarthy Group Tools (<https://www.chg.ox.ac.uk/~wrayner/tools/>) with parameters -a -h -n, given the BIM file of the PLINK dataset, the .frq file generated by --freq, and the TopMed reference panel. After compression with bgzip, the VCF files were uploaded to the TOPMed Imputation Server,¹⁵⁻¹⁷ where genotype phasing and imputation was performed using the TOPMed r3 imputation panel, array build = GRCh38/hg38, rsq Filter = 0.3, Eagle v2.4 phasing, population = "vs. TOPMed Panel", and mode = "Quality Control and Imputation". Finally, the imputed per-chromosome VCF files, containing 445,271,096 variants in total, were filtered by bcftools filter -i 'INFO/R2>=0.7' (bcftools v1.17), retaining 62,322,707 variants in total.

HLA imputation

First, the reference allele of each SNP was identified from the array's SNP annotation file. Then, a VCF file was created over the remaining SNPs in chr6:28002253-33997776 (the region covered by the reference panel used) that had a reference allele, and the remaining samples. The file was created by --recode vcf-iiid in PLINK 1.9, using --a2-allele to force the A2 allele to the SNP reference allele and --real-ref-alleles to also remove PR from the INFO values. After compression with bgzip, the file was uploaded to the Michigan Imputation Server,¹⁶ where genotype phasing and imputation was performed using the Four-digit Multi-ethnic HLA reference panel v2, array build = GRCh38/hg38, Eagle v2.4 phasing, and mode = "Quality Control and Imputation". The imputed VCF file, containing 22,734 variants, was filtered by bcftools filter -i 'INFO/R2>=0.7', resulting in a file with 21,303 variants.

Generation of genome-wide and extended chromosome-6 datasets

A VCF file containing the imputed intergenic SNPs (variants with names starting with "rs") was created from the imputed HLA VCF file and lifted over to hg38 using Picard's LiftoverVcf tool. SNPs present in both that file and the imputed chromosome-6 VCF file were then identified and removed from a file if their INFO/R2 was higher in the other file. The per-chromosome imputed VCF files, excluding chromosome 6, the filtered intergenic-HLA-SNP VCF file, and the filtered chromosome-6 imputed VCF file were then combined and converted to a GDS file by the seqVCF2GDS function in the SeqArray R package¹⁸ (v1.40.1) with scenario = "imputation" and reference = "hg38", in order to be used in R. The file contained 62,323,890 variants, including 163,125 variants in the xMHC. We refer to this dataset as the genome-wide dataset.

A VCF file containing the remaining (non-intergenic-SNP) imputed HLA variants (which we refer to as extended HLA variants) was created from the imputed HLA VCF file. The position of each HLA allele was set to the mid hg38 position of the corresponding HLA gene, while the position of each other variant was set to the position extracted from its name, after lifting it over to hg38. Finally, the filtered chromosome-6 imputed VCF file, the filtered intergenic-SNP HLA VCF file, and the non-intergenic-SNP HLA VCF file were combined and

converted to a GDS file by `SeqArray::seqVCF2GDS()`. The file contained 3,629,020 variants, including 169,753 variants in the xMHC. We refer to this dataset as the extended chromosome-6 dataset.

Pan-ancestry association analysis

AQP4-positive NMOSD cases and controls were first selected. Variants with $MAF \geq 0.05$ were then selected using the `alleleFrequency` function in the `SeqVarTools` R package (v1.38.0) with `genome.build = "hg38"` given the sample sex. A null model was learned from the 12 PCs, sex, and the final set of kinship coefficients (multiplied by 2) by `GENESIS::fitNullModel()` with `family = "binomial"`. Association analysis was performed by `GENESIS::assocTestSingle()` with `test = "Score.SPA"`, `imputed = TRUE`, and `genome.build = "hg38"`, given the null model.

Ancestry assignment

The ancestry of each sample was initialised with its assigned 1KG superpopulation. However, some ethnic backgrounds, including mixed backgrounds, are not represented in 1KG. The distance to the nearest 1KG sample would be expected to be larger for samples from those backgrounds compared to the other samples. The distance distribution was indeed bimodal (supplementary figure 7). A threshold that separated the two modes was identified by log-transforming the distance, discretising log-distance into 100 equal-sized bins, thresholding the discrete distribution using the `auto_thresh` function in the `autothresholdr` R package (v1.4.1) with `method = "Minimum"`, converting the threshold back to the continuous scale, and, finally, exponentiating the threshold. Samples with distance beyond that threshold were considered ancestry outliers and their ancestry was set to missing. 78 (91.8%) of the 85 of Greater Middle Eastern samples were indeed outliers (supplementary table 4). 77 of the Greater Middle Eastern samples were from Iran, and 76 (98.7%) of those were outliers. 21 (28.0%) out of the 75 samples of mixed background were also outliers. Finally, the ancestry of samples with ethnic background American Indian or Alaska native, Greater Middle Eastern, or Native Hawaiian/Pacific Islander was set to their ethnic background regardless of outlier status; outliers of other backgrounds were considered to be of mixed ancestry and were not used in ancestry-specific analyses.

Principal component analysis for Europeans

A set of unrelated samples (UNRELATED_EUR) was identified among the remaining European samples using `PC_RELATE_2`, and SNPs with $MAF \geq 0.05$ in that set were identified. Among those SNPs, SNPs in HWE were identified using `RUTH` with `UNRELATED_3` and `PCS_2` and LD-pruned using `UNRELATED_EUR`. A PC transformation over the LD-pruned SNPs with 20 PCs was learned from `UNRELATED_EUR`. The number of PCs to retain was determined to be 4, with PC4 associated with Sweden. The PC transformation with 4 PCs was applied to all remaining European samples to generate `PCS_EUR`.

Identification of independent associations

The (non-symmetric) MMPC algorithm¹⁹ was employed to identify independent associations (supplementary algorithm 1). MMPC can be viewed as a feature selection algorithm with a forward selection and a backward elimination phase. MMPC works by performing tests of conditional independence, which, in our case, are tests of independence of a variant X from the phenotype P given a subset Z of the rest variants. Suppose that X is marginally associated with P and that Z , but no proper subset of it, renders X independent from P . According to the necessary path condition,²⁰ every variant in Z is marginally associated with X . The principle of independent assortment then implies that all variants in Z are on the same chromosome as X . For this reason, MMPC was applied separately to each chromosome and the results were combined. The significance level of the tests was set to 10^{-5} . Owing to the observation that LD decays considerably within 500 kb in humans (e.g., see Figure 3 in Ke and colleagues),²¹ tests of independence of a variant X from P given a subset Z of the rest variants on the same chromosome were restricted to Z in which every variant is within 1 mb of X . Due to the existence of long-range LD throughout the xMHC, the 1 mb threshold was not used within the xMHC. Specifically, tests of independence of an xMHC variant X from P given a subset Z of the rest variants on the same chromosome were restricted to Z in which every variant is also in the xMHC or within 1 mb of X .

Our own R implementation of MMPC was used, which defines an interface for tests of independence of a variable X from each of the variables in a set Y given a set Z . By testing all variables in Y for a certain Z in the same call, the null model of likelihood-ratio tests like the one in `GENESIS` is learned only once for each Z . The `fitNullModel` and `assocTestSingle` functions in `GENESIS` were wrapped into an implementation of the interface to perform tests of independence of a variant X from P given a subset Z of the rest variants, the PCs, and sex.

For empty Z , the test reduces to that used in the association analysis; the test was therefore wrapped in a version that reused those p-values. Our implementation of MMPC also uses an interface to decide which variables in a set Y can be rendered independent from a variable X given a set of variables Z . An implementation of the interface was used to decide which variants in a set Y can be rendered independent from the phenotype P given a set of variants Z . The implementation selects an xMHC variant Y if every variant in Z is also in the xMHC or within 1 mb of Y , and a non-xMHC variant Y if every variant in Z is within 1 mb of Y .

Indirect matching of independent associations

For an unordered pair of studies $\{S1, S2\}$ (either ours or in the GWAS Catalog), independent association $I1$ of variant $V1$ in $S1$ was matched to independent association $I2$ of a different variant $V2$ in $S2$ if $V1$ and $V2$ are proxies for each other (within 1 mb of each other and with $r^2 \geq 0.7$) in a population (1KG superpopulation) P included in both studies (for example, all 1KG superpopulations are included in our pan-ancestry study, while only EUR is included in our European study). If $S1$ was specific to P , it was additionally required that $V2$ was associated ($p < 10^{-5}$) with the phenotype in $S1$, if $V2$ was present in $S1$. This association is expected if either variant is causal or they are both proxies for the same causal variant, while its absence would suggest that the variants are proxies for different causal variants. Accordingly, if $S2$ was specific to P , it was additionally required that $V1$ was associated with the phenotype in $S2$, if $V1$ was present in $S2$. To determine whether a variant was present and associated with the phenotype in a GWAS-Catalog study, its harmonized summary statistics were downloaded, if available; if the statistics were not available, the condition was not checked.

Variant proxies were identified from two sources, referred to as “LD sources”. The first LD source was LDproxy.²² The coordinates of the variant were supplied to the LDproxy function of the LDlinkR R package²³ (v1.4.0) with genome build = “grch38” and win_size = 1000000. After checking whether the alleles of the first proxy returned match those of the variant, the first proxy and proxies with $r^2 < 0.7$ were removed. Indels were converted between the left-normalised representation in our dataset and the dash-based representation in LDproxy. The second LD source was the xMHC portion of our European extended chromosome-6 dataset, which was used to find proxies throughout the xMHC in Europeans without a distance limitation. A set of unrelated samples (UNRELATED_EUR_CONTROLS) was first identified among the remaining European controls using PC_RELATE_2. The proxies for an xMHC variant were identified by calculating its partial correlation with each other xMHC variant given PCS_EUR and sex using psych::partial.r(), squaring the result, and removing proxies with $r^2 < 0.7$.

Meta-analysis with the study of Estrada and colleagues

To make additional discoveries in Europeans, we sought to meta-analyse our European study with that of Estrada and colleagues.²⁴ Estrada and colleagues stated that their cases may represent up to 23% of all NMO cases in the USA. Since a large portion of our cases came from the USA and AQP4-positive NMOSD is a rare disease, we assumed that there is sample overlap between the studies. Thus, we opted to use sample size weighted meta-analysis in METAL²⁵ (release 2020-05-05), which supports correction for sample overlap. As our study contained related samples, however, the sample size could not be used.²⁶ An association analysis was therefore conducted by --glm omit-ref sex in PLINK 2, with the PCs given as covariates with --covar, using the set of unrelated samples previously identified for fine-mapping (1,741 samples, comprising 767 cases and 974 controls) and the 6,001,273 variants with MAF ≥ 0.05 in that set. Harmonised summary statistics for the study of Estrada and colleagues were downloaded from the GWAS Catalog. The summary statistics in each study of the 5,141,744 variants present in both studies were then supplied to METAL to conduct a sample size weighted meta-analysis with correction for sample overlap. METAL estimated that 14.8% of the samples of Estrada and colleagues overlap with ours.

To identify independent associations, MMPC could have been used with GCTA-COJO²⁷ to perform approximate conditional tests given summary statistics and an external LD reference panel. GCTA-COJO, however, requires effect sizes, and those are not generated by sample size weighted meta-analyses in METAL. Supplementary algorithm 2 was therefore used to identify lead variants in regions of 1 mb or a superset of the xMHC (expanded up to 1 mb on one side) among the associated variants instead.

Bayesian fine-mapping

A genomic region was created around each independently associated variant by expanding its genomic position 1 mb on each side. The region around each independently associated xMHC variant was merged with the entire xMHC, due to the existence of long-range LD throughout the latter. Overlapping regions were merged. A set of

unrelated samples was identified among the cases and controls using PC_RELATE_2. For each fine-mapping region, partial correlations between each pair of variants in the region conditional on the PCs and sex were computed by the partial.r function in the psych R package (v2.4.3) using those samples (for the European study, the European samples and PCs were used). Bayesian fine-mapping of the region was then performed by the summary-statistics version of SuSiE^{28,29} (susieR R package v0.12.35) with the variant z-scores, variant correlations, and sample size as input.

Bayesian colocalisation of GWAS and eQTL signals

First, the studies in the eQTL Catalogue³⁰ were either assigned a 1KG superpopulation or marked as multi-ancestry as follows. The number of samples of each study that were assigned to each 1KG superpopulation was obtained from https://github.com/eQTL-Catalogue/eQTL-Catalogue-resources/blob/master/data_tables/population_assignments.tsv. Studies with $\geq 90\%$ of the samples assigned to a single population were assigned to that population, while the remaining studies were marked as multi-ancestry. Studies not in that file were assigned to a single population or marked as multi-ancestry after consulting the corresponding publications.

For each European dataset in the eQTL Catalogue, colocalization was then tested between each fine-mapped region of our European study and overlapping fine-mapped regions (corresponding to molecular traits) in the eQTL-Catalogue dataset using the coloc R package³¹ (v5.2.3) with the SuSiE-computed log Bayes factors (LBFs) of each region as input.

Cell type and tissue enrichment

Enrichment of disease heritability in cell types and tissues (corresponding to GTEx datasets) was tested by S-LDSC^{32,33} using the summary statistics from our European study, baseline model v1.2 (1000G_Phase3_baseline_v1.2_ldscores, downloaded from <https://zenodo.org/records/10515792>), the latest version of the LD scores for the GTEx datasets supplied to us by the Broad Institute (Konrad J Karczewski, Benjamin M Neale), and regression weights with (1000G_Phase3_ldscores) and without (1000G_Phase3_weights_hm3_no_MHC) the MHC. There were no significant results after FDR correction in either case (appendix 2 pp 24-25).

Genetic risk sharing with other autoimmune diseases

GWAS-Catalog studies on autoimmune diseases (other than AQP4-positive NMOSD) in Europeans with independent associations matched to those in our European study were first identified, and studies of subclasses of the same disease as another study were dropped. After adding a GWAS-Catalog study on multiple sclerosis in Europeans³⁴ and our European study to the studies, independent associations were matched across the studies. The matching coefficient between two studies was defined as the weighted mean proportion of matched independent associations, the weight of each study being the number of independent associations. For studies on the same disease, the study with the greatest matching coefficient with ours was selected (supplementary table 7) and a graph of diseases was plotted (figure 3). After removing GWAS-Catalog studies without harmonised summary statistics available, resulting in a one-to-one correspondence between studies and diseases, the genetic correlation between each pair of diseases was tested by LDSC^{32,35} (supplementary table 8 and supplementary figure 9). Finally, a literature review was performed to determine the efficacy of currently available non-steroidal treatments for multiple sclerosis, NMOSD, and a selection of other autoimmune diseases. The full list of references used to classify the efficacy of each treatment in each disease is available in appendix 2 p 23. We acknowledge that the observations by clinicians for some of these drugs and associated treatment efficacy for a specific disease may differ from those reported in the literature.

Supplementary notes

Supplementary note 1

The reason for using the entire chromosome 6 and selecting the identified independent associations within the xMHC instead of using the xMHC directly is the following. Suppose that causal variant X on chromosome 6 is outside the xMHC but in LD with variant Y in the xMHC. If the entire chromosome 6 is used, X will be identified as an independently associated variant and Y will not. If the xMHC is analysed in isolation, however,

Y will be identified (under ideal conditions) as an independently associated variant. This is undesirable, as we do not want to capture signals outside the xMHC in the comprehensive xMHC analysis.

Supplementary note 2

To further explore the relationship between HLA-DRB1 alleles and amino acids, we calculated the frequency in our European controls of each allele/amino-acid pair for positions 74 (appendix 2 p 26) and 77 (appendix 2 p 27). For position 74, having HLA-DRB1*03(:01) implied having R at that position, which was also observed with HLA-DRB1*01. HLA-DRB1*01:01, HLA-DRB1*04:01, HLA-DRB1*11, HLA-DRB1*13, and HLA-DRB1*15(:01) all implied A, which was protective ($p = 2.0872 \times 10^{-16}$, OR = 0.53) and was also observed with HLA-DRB1*01 and HLA-DRB1*04. For position 77, HLA-DRB1*03(:01) implied N, which was also observed for HLA-DRB1*01. HLA-DRB1*01:01, HLA-DRB1*04:01, HLA-DRB1*07:01, HLA-DRB1*11, HLA-DRB1*13, and HLA-DRB1*15(:01) all implied T, which was protective ($p = 2.5936 \times 10^{-26}$, OR = 0.36) and was also observed with HLA-DRB1*01, HLA-DRB1*04, and HLA-DRB1*07.

Supplementary note 3

It is not uncommon for a GWAS to exclude the (x)MHC, which might have resulted in fewer matches than possible with AQP4-positive NMOSD. The MHC was indeed excluded from the most recent study of multiple sclerosis in Europeans.³⁶ In the most recent multi-ancestry study of multiple sclerosis,³⁴ a genome-wide association analysis excluding the xMHC was performed, whose results were uploaded to the GWAS Catalog, while the xMHC was analysed separately using all types of HLA variation. It is possible that the (x)MHC was excluded from all other multiple sclerosis studies in the GWAS Catalog as well, but this was not checked.

The International NMOSD Genetics Consortium

List of contributing authors who collected and contributed DNA samples from NMOSD patients and controls for processing and analysis in our AQP4-positive NMOSD GWAS study.

First name	Last name	Affiliation
Orhan	Aktas	Düsseldorf MS & NMOSD & MOGAD center Department of Neurology, Heinrich Heine University Düsseldorf
Nasrin	Asgari	Institute of Regional Health Research & Institute of Molecular Medicine, University of Southern Denmark
Ilya	Ayzenberg	Klinik für Neurologie, Ruhr Universität Bochum
Jacinta	Behne	The Guthy-Jackson Charitable Foundation
Judith	Bellmann-Strobl	Charité – Universitätsmedizin Berlin
Thomas	Berger	Dept. of Neurology, Medical University of Vienna, Vienna, Austria, Comprehensive Center for Clinical Neurosciences & Mental Health, Medical University of Vienna, Vienna, Austria
Achim	Berthele	Department of Neurology, School of Medicine, Technical University Munich, Klinikum Rechts der Isar, Munich, Germany.
Yolanda	Blanco	Neuroimmunology and Multiple Sclerosis Unit, Laboratory of Advanced Imaging in Neuroimmunological Diseases, Hospital Clinic Barcelona, Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS) and Universitat de Barcelona, 08036 Barcelona, Spain.
Lou	Brundin	Department of Clinical Neuroscience, Division of Neurology, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden.
Manuel	Comabella-López	Servei de Neurologia. Centre d'Esclerosi Múltiple de Catalunya (Cemcat). Institut de Recerca Vall d'Hebron (VHIR). Hospital Universitari Vall d'Hebron. Universitat Autònoma de Barcelona. Barcelona, Spain

Larry	Cook	University of Utah, Department of Pediatrics, Division of Critical Care
Jelena	Drulovic	Clinic of Neurology, Faculty of Medicine, University of Belgrade
Irena	Dujmovic Basuroski	University of Belgrade School of Medicine, Clinical Centre of Serbia, Department of Neurology, Belgrade, Serbia and University of North Carolina School of Medicine, Department of Neurology, Chapel Hill, NC, USA.
Maj Lisa	Frankenberg	Department of neurology, University Hospital Münster, Münster, Germany
Katrin	Giglhuber	Department of Neurology, School of Medicine and Health, Technical University Munich, Klinikum rechts der Isar, Munich, Germany
Maria	Hastermann	Experimental and Clinical Research Center, Max Delbrueck Center for Molecular Medicine and Charité Universitätsmedizin Berlin, Berlin, Germany
Jyh Yung	Hor	Department of Neurology, Penang General Hospital, Penang, Malaysia
Saif	Huda	Walton Centre NHS Foundation Trust, University of Liverpool, UK
Martin W.	Hümmert	Department of Neurology, Hannover Medical School, Germany.
Ellen	Iacobaeus	Department of Clinical Neuroscience, Karolinska Institutet and Department of Neurology, Karolinska University Hospital, SE-171 76 Stockholm, Sweden
Dagur	Inge Jonsson	The Karolinska Neuroimmunology and Multiple Sclerosis Centre, Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden
Raffaele	Iorio	Department of Neuroscience, Università Cattolica del Sacro Cuore, Rome Italy and Neurology Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy
Noriko	Isobe	Department of Neurology, Graduate School of Medical Sciences, Kyushu University
Anu	Jacob	Cleveland Clinic Abu Dhabi and Walton Centre NHS Foundation Trust, University of Liverpool, UK
Sven	Jarius	Molecular Neuroimmunology Group, Department of Neurology, University of Heidelberg, Heidelberg, Germany
Megan	Kenneally	The Guthy-Jackson Charitable Foundation
Jun-ichi	Kira	Department of Neurology, Kyushu University, Fukuoka, Japan
Mohsen	Khademi	Department of Clinical Neuroscience, Karolinska Institutet and Department of Neurology, Karolinska University Hospital, SE-171 76 Stockholm, Sweden
Luisa	Klotz	Department of neurology, University Hospital Münster, Münster, Germany
Tania	Kümpfel	Institute of Clinical Neuroimmunology, University Hospital and Biomedical Center, Ludwig-Maximilians University Munich, Germany
Romain	Marignier	Centre de Référence des Maladies Inflammatoires Rares du Cerveau et de la Moelle, Hôpital Neurologique Pierre Wertheimer, Hospices Civils de Lyon, Lyon, France and Department of Neurology, Sclérose en Plaques, Pathologies de la Myéline et Neuro-Inflammation, Lyon, France.
Marcelo	Matiello	Department of Neurology, Massachusetts General Brigham, Harvard Medical School, Boston, MA, USA
Janine	Meyer	Department of neurology, University Hospital Münster, Münster, Germany

Xavier	Montalban	Servei de Neurologia. Centre d'Esclerosi Múltiple de Catalunya (Cemcat). Institut de Recerca Vall d'Hebron (VHIR). Hospital Universitari Vall d'Hebron. Universitat Autònoma de Barcelona. Barcelona, Spain
Been	Olafur Sveinsson	The Karolinska Neuroimmunology and Multiple Sclerosis Centre, Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden
Tomas	Olsson	The Karolinska Neuroimmunology and Multiple Sclerosis Centre, Department of Clinical Neuroscience, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden
Friedemann	Paul	Experimental and Clinical Research Center, Max Delbrueck Center for Molecular Medicine and Charité Universitätsmedizin Berlin, Berlin, Germany
Markus	Reindl	Clinical Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria
Lisa	Revie	Department of neurology, University Hospital Münster, Münster, Germany
Marius	Ringelstein	Department of Neurology, Medical Faculty and University Hospital, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany and Department of Neurology, Center for Neurology and Neuropsychiatry, LVR-Klinikum, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany
Neil	Robertson	Department of Neurology, Institute of Psychological Medicine and Clinical Neuroscience, Cardiff University University Hospital of Wales, Heath Park Cardiff, UK
Paulus	Rommer	Department of Neurology, Medical University of Vienna, Vienna 1090, Austria and Comprehensive Center for Clinical Neurosciences and Mental Health, Medical University of Vienna, Vienna 1090, Austria
Mohamad Ali	Sahraian	Multiple Sclerosis Research Center, Neuroscience Institute, Tehran University of Medical Sciences
Ayako	Sakoda	Department of Neurology, Neurological Institute, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
Ernestina	Santos	Neurology Department, Hospital Santo António, Largo Professor Abel Salazar, 4099-001 Porto
Chanjira	Satukijchai	Brain Center, Bangkok Hospital Headquarters, Thailand, 10310
Maria	Sepúlveda	Neuroimmunology and Multiple Sclerosis Unit and Laboratory of Advanced Imaging in Neuroimmunological Diseases (ImaginEM), Hospital Clinic and Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Spain
Sasitorn	Siritho	Siriraj Neuroimmunology Center, Siriraj Hospital, Mahidol University, Thailand, 10700
Keiko	Tanaka	Department of Neurology, Kanazawa Medical University, Kanazawa Japan (currently: Department of Animal Model Development, Brain Research Institute, Niigata University, Niigata, Japan)
Masami	Tanaka	Multiple Sclerosis Center, Utano National Hospital, National Hospital Organization, Kyoto, Japan. (currently: Department of Neurology, Kyoto MS Center, Kyoto Min-Iren Chuo-Hospital, Kyoto, Japan)
Pentti J	Tienari	Department of Neurology, Brain Center, Helsinki University Hospital, Helsinki, Finland and Translational Immunology, Research Programs Unit, University of Helsinki, Helsinki, Finland.

Corinna	Trebst	Department of Neurology, Hannover Medical School, Germany.
Andrés María	Villa	Sección Neuroinmunología. Hospital Ramos Mejía. Facultad de Medicina. Universidad de Buenos Aires
Angela	Vincent	Nuffield Department of Clinical Neurology, University of Oxford, UK
Mitsuru	Watanabe	Department of Neurology, Kyushu University Hospital. 3-1-1 Maidashi, Higashi-ku, Fukuoka, 812-8582, Japan
Brigitte	Wildemann	Molecular Neuroimmunology Group, Department of Neurology, University of Heidelberg, Heidelberg, Germany
Michael	Yeaman	The Guthy-Jackson Charitable Foundation
Alexander	Zimprich	Medical University of Vienna, Vienna 1090, Austria

Description of Appendix 2 pages

Page 1

Independently associated variants for the pan-ancestry study.

Column(s)	Description
name	Name in the form chr:pos:ref:alt.
rs	RS ID; blank if unavailable.
chr, pos, ref, alt	Chromosome, position (hg38), reference allele, and alternative allele, respectively.
imputation_status	Imputation status, either: <ul style="list-style-type: none"> • imputed: imputed and not typed, • typed: typed and imputed, or • typed_only: typed and not imputed.
r2	Imputation score.
ref_freq, alt_freq	Frequency of the reference and alternative allele, respectively.
p_value	Association p-value.
or	Association odds ratio.
ll	Lower limit of the 95% confidence interval for or.
ul	Upper limit of the 95% confidence interval for or.
cond_p_value	Maximal p-value among the conditional tests performed for the variant.
cond_or	Odds ratio corresponding to cond_p_value.
cond_set	Conditioning set corresponding to cond_p_value.
region	Fine-mapped region in the form chr:start:end.
pip	Posterior inclusion probability (PIP).
credible_set	The fine-mapped region's 95% credible set where the variant is included; blank if the variant is not included in a credible set.
rank_in_credible_set	The rank of the variant in its credible set; the rank of tied variants is their mean rank; blank if the variant is not included in a credible set.
associated_genes	Genes whose expression is associated with the variant in some dataset in the eQTL Catalogue (comma-separated).

Rows are sorted by p_value.

Page 2

Variants in credible sets for the pan-ancestry study.

Column(s)	Description
region	Fine-mapped region in the form chr:start:pos.
credible_set	Credible set of that region.
rank	Rank of the variant in the credible set.
variant	Name of the variant in the form chr:pos:ref:alt.
rs	RS ID of the variant; blank if unavailable.

p_value	Association p-value.
or	Association odds ratio.
ll	Lower limit of the 95% confidence interval for or.
ul	Upper limit of the 95% confidence interval for or.
pip	Posterior inclusion probability (PIP).

Rows are sorted by region, credible_set, and rank.

Page 3

Associations between independently associated variants in our pan-ancestry study and molecular traits in eQTL-Catalogue datasets. Each row is a combination (X, M, D) of

- an independently associated variant X in our study (S_1),
- a molecular trait M , and
- an eQTL-Catalogue dataset D where X is associated with M .

D itself is a combination (S_2, G, Q) of

- an eQTL-Catalogue study S_2 ,
- a sample group G , corresponding to a cell type / tissue and condition in S_2 , and
- a quantification method Q .

Column(s)	Description
variant	Name of X in the form chr:pos:ref:alt.
rs	RS ID of X ; blank if unavailable.
p_value	Association p-value of X in S_1 .
or	Association odds ratio of X in S_1 .
molecular_trait_id	ID of M . Depends on Q .
gene_id	Ensembl gene id of M .
gene_symbol	HGNC symbol of M ; blank if unavailable.
dataset	ID of D .
study	Label of the study (S_2) of D .
type	Type of S_2 : bulk_rna_seq, microarray, single_cell_rna_seq, or protein_qtl.
population	1KG superpopulation of S_2 , blank in case of a multi-ancestry study.
sample_group	Sample group (G) of D .
quant_method	Quantification method (Q) of D .
dataset_p_value	Association p-value of X in D .
dataset_beta	Association beta coefficient of X in D .

Rows are sorted by p_value and dataset_p_value.

Page 4

Independently associated variants for the European study. In the same format as page 1.

Page 5

Directly matched independent associations between the studies (S_1 = European study, S_2 = pan-ancestry study). Each row corresponds to a variant that is an independently associated variant in both studies.

Column(s)	Description
name	Name in the form chr:pos:ref:alt.
rs	RS ID; blank if unavailable.
p_value_1	Association p-value in S_1 .
p_value_2	Association p-value in S_2 .
or_1	Association odds ratio in S_1 .
or_2	Association odds ratio in S_2 .

Page 6

Indirectly matched independent associations between the studies (S_1 = European study, S_2 = pan-ancestry study). Each row is a combination (X_1, L, P, X_2) of

- an independently associated variant X_1 in S_1 ,
- a population P (EUR when either S_1 or S_2 is European and any 1KG superpopulation when both S_1 and S_2 are pan-ancestry),
- an LD source L (either LDProxy or our xMHC dataset), and
- an independently associated variant X_2 in S_2 that is a proxy for X_1 in P using L .

Column(s)	Description
variant_1	Name of X_1 in the form chr:pos:ref:alt.
rs_1	RS ID of X_1 ; blank if unavailable.
variant_2	Name of X_2 in the form chr:pos:ref:alt.
rs_2	RS ID of X_2 ; blank if unavailable.
distance	Distance between X_1 and X_2 .
population	Population where X_1 and X_2 are proxies for each other (P). A 1KG superpopulation.
source	LD source (L). Either ld_proxy (LDProxy) or xmhc (our xMHC dataset).
r2	Strength of the LD (r^2) between X_1 and X_2 in P using L .
ref_1_allele_2, alt_1_allele_2	Allele of X_2 that is correlated with the reference and the alternative allele of X_1 , respectively.
variant_1_p_value_1	Association p-value of X_1 in S_1 .
variant_1_p_value_2	Association p-value of X_1 in S_2 .
variant_2_p_value_1	Association p-value of X_2 in S_1 .
variant_2_p_value_2	Association p-value of X_2 in S_2 .
or_1_p_value_1	Association odds ratio of X_1 in S_1 .
or_1_p_value_2	Association odds ratio of X_1 in S_2 .
or_2_p_value_1	Association odds ratio of X_2 in S_1 .
or_2_p_value_2	Association odds ratio of X_2 in S_2 .

Page 7

Lead associations in the meta-analysis. In the same format as page 1, but with a subset of the columns.

Page 8

Directly matched independent associations between the meta-analysis (S_1) and the pan-ancestry study (S_2). In the same format as page 5, but odds ratios are missing for the meta-analysis.

Page 9

Indirectly matched independent associations between the meta-analysis (S_1) and the pan-ancestry (S_2). In the same format as page 6, but odds ratios are missing for the meta-analysis.

Page 10

Directly matched independent associations between the European study (S_1) and the meta-analysis (S_2). In the same format as page 5, but odds ratios are missing for the meta-analysis.

Page 11

Indirectly matched independent associations between the European study (S_1) and the meta-analysis (S_2). In the same format as page 6, but odds ratios are missing for the meta-analysis.

Page 12

Variants in credible sets for the European study. In the same format as page 2.

Page 13

Associations between independently associated variants in our European study and molecular traits in eQTL-Catalogue datasets. In the same format as page 3.

Page 14

Colocalisation results. Each row is a pair (X, Y) of

- a credible set X from our European study, and
- a credible set Y from a European dataset in the eQTL Catalogue

for which colocalization analysis was performed (i.e., the corresponding fine-mapped regions overlapped, and there were sufficient variants shared). X itself is a combination (R, I_1) of

- a fine-mapped region R , and
- a credible-set index I_1 .

D itself is a combination (S, G, Q, M, I_2) of

- an eQTL-Catalogue study S ,
- a sample group G , corresponding to a cell type / tissue and condition in S ,
- a quantification method Q ,
- a molecular trait M , and
- a credible-set index I_2 .

Column(s)	Description
region	Name of R in the form chr:start:pos.
dataset	ID of D .
study	Label of the study (S) of D .
type	Type of S : bulk_rna_seq, microarray, single_cell_rna_seq, or protein_qtl.
sample_group	Sample group (G) of D .
quant_method	Quantification method (Q) of D .
molecular_trait_id	ID of M . Depends on Q .
gene_id	Ensembl gene id of M .
gene_symbol	HGNC symbol of M ; blank if unavailable.
credible_set	I_1
dataset_credible_set	I_2
pp_0	Probability of no causal variant in either credible set.
pp_1	Probability of causal variant in X only.
pp_2	Probability of causal variant in Y only.
pp_3	Probability of two distinct causal variants.
pp_4	Probability of one common causal variant.
top_variant	Name, in the form chr:pos:ref:alt, of the variant T with the highest probability of being the one common causal variant, if there is one common causal variant.
top_variant_rs	RS ID of T ; blank if unavailable.
top_variant_pp	Probability of T being the one common causal variant, if there is one common causal variant.

Probabilities pp_0 to pp_4 sum to 1. Rows are sorted by pp_4.

Page 15

Independently associated xMHC variants for the pan-ancestry comprehensive xMHC analysis. In the same format as page 1, except that there are no columns chr and associated_genes. Intergenic-SNP xMHC variants are named as above. Non-intergenic-SNP HLA variants (extended HLA variants) are named as described in <https://github.com/immunogenomics/HLA-TAPAS/wiki/HLA-imputation>. The position of classical HLA alleles is the mid position of the HLA gene.

Page 16

Independently associated xMHC variants for the European xMHC analysis. In the same format as page 1, except that there are no columns chr and associated_genes. Intergenic-SNP xMHC variants are named as above. Non-intergenic-SNP HLA variants (extended HLA variants) are named as described in <https://github.com/immunogenomics/HLA-TAPAS/wiki/HLA-imputation>. The position of classical HLA alleles is the mid position of the HLA gene.

Page 17

Association results for each xMHC variant upon conditioning on the top i independently associated HLA variants ($0 \leq i \leq n$, where n is the number of independently associated xMHC variants) for the pan-ancestry comprehensive xMHC analysis.

Column(s)	Description
name	Name in the form chr:pos:ref:alt.
rs	RS ID; blank if unavailable.
p_value_ i	Association p-value upon conditioning on the top i independently associated xMHC variants. For $i = 0$, it is the marginal p-value. Blank for the k -th top independently associated xMHC variant for $i \geq k$.
or_ i	Odds ratio corresponding to p_value_ i .
ll_ i	Lower limit of the 95% confidence interval for or_ i .
ul_ i	Upper limit of the 95% confidence interval for or_ i .

Page 18

Association results for each xMHC variant upon conditioning on the top i independently associated HLA variants ($0 \leq i \leq n$, where n is the number of independently associated xMHC variants) for the European comprehensive xMHC analysis. In the same format as page 17.

Page 19

Direct matches between independent associations in our pan-ancestry study and independent associations in GWAS-Catalog studies. Each row is a pair (X , S_2) of an independently associated variant X in our study (S_1) and a GWAS-Catalog study S_2 where X is also an independently associated variant.

Column(s)	Description
variant	Name of X in the form chr:pos:ref:alt.
rs	RS ID of X ; blank if unavailable.
p_value	Association p-value of X in S_1 .
or	Association odds ratio of X in S_2 .
study	Accession of S_2 .
pubmed_id	PubMed ID of S_2 .
disease_or_trait	Name of the disease or trait of S_2 .
first_author	Name of the first author of S_2 .
title	Title of the publication of S_2 .
journal	Name of the journal where S_2 was published.
date	Date S_2 was published.
ancestries	Comma-separated list of broad ancestral categories across the samples in the initial stage of S_2 .
study_risk_allele	Risk allele of X in S_2 ; blank if unavailable. Might not match ref or alt. ¹
study_p_value	Association p-value of X in S_2 .
study_p_value_text	Context of study_p_value.
study_or_or_beta	Association odds ratio or beta coefficient of X in S_2 .

Rows are sorted by p_value and study_p_value.

¹ According to the GWAS Catalog FAQs (accessed June 28, 2024): “Variants and their risk alleles are extracted exactly as reported in the paper. In a small number of cases, the curated risk allele does not match either of the alleles reported in the reference genome assembly, which are displayed in the Info panel on the variant page. This may be due to strand flipping between genome builds, where the association was originally reported on an older genome build. We are unable to account for this in the curated data as genome build is not consistently reported by authors.”

Page 20

Direct matches between independent associations in our European study and independent associations in GWAS-Catalog studies. In the same format as page 19.

Page 21

Indirect matches between independent associations in our pan-ancestry study and independent associations in GWAS-Catalog studies. Each row is a combination (X_1, S_2, P, L, X_2) of

- an independently associated variant X_1 in our study (S_1),
- a GWAS-Catalog study S_2 ,
- a population P (a 1KG superpopulation that was included in either S_1 or S_2),
- an LD source L (either LDProxy or our xMHC dataset), and
- an independently associated variant X_2 in S_2 that is a proxy for X_1 in P using LD source L .

Column(s)	Description
variant	Name of X_1 in the form chr:pos:ref:alt.
rs	RS ID of X_1 ; blank if unavailable.
p_value	Association p-value of X_1 in S_1 .
or	Association odds ratio of X_1 in S_1 .
proxy	Name of X_2 in the form chr:pos:ref:alt.
proxy_rs	RS ID of X_2 ; blank if unavailable.
distance	Distance between X_1 and X_2 .
population	1KG superpopulation where X_1 and X_2 are proxies for each other (P).
source	Source of r^2 (L). Either ld_proxy (LDProxy) or xmhc (our xMHC dataset).
r2	Strength of the LD (r^2) between X_1 and X_2 in P using L .
variant_ref_cor_allele, variant_alt_cor_allele	Allele of X_2 that is correlated with the reference and the alternative allele of X_1 , respectively.
study	Accession of S_2 .
pubmed_id	PubMed ID of S_2 .
disease_or_trait	Name of the disease or trait of S_2 .
first_author	Name of the first author of S_2 .
title	Title of the publication of S_2 .
journal	Name of the journal where S_2 was published.
date	Date S_2 was published.
ancestries	Comma-separated list of broad ancestral categories across the samples in the initial stage of S_2 .
study_risk_allele	Risk allele of X_2 in S_2 ; blank if unavailable. Might not match ref or alt.
study_p_value	Association p-value of X_2 in S_2 .
study_p_value_text	Context of study_p_value.
study_or_or_beta	Association odds ratio or beta coefficient of X_2 in S_2 .

Rows are sorted by p_value and study_p_value.

Page 22

Indirect matches between independent associations in our European study and independent associations in GWAS-Catalog studies. In the same format as page 21.

Page 23

References for each treatment and disease pair. Rows are treatments and columns are diseases.

Page 24

Results of cell-type/tissue enrichment tests with the MHC. Each row is a GTEx dataset corresponding to a cell type or tissue.

Column(s)	Description
-----------	-------------

name	Dataset name.
beta	Correlation coefficient.
p_value	P-value.
adj_p_value	FDR-adjusted p-value

Rows are sorted by adj_p_value (or, equivalently, p_value).

Page 25

Results of cell-type/tissue enrichment tests without the MHC. In the same format as page 24.

Page 26

Frequency of combinations of amino acids at position 74 of the HLA-DRB1 protein and classical *HLA-DRB1* alleles. Rows are amino acids (or amino-acid combinations) at that position and columns are classical *HLA-DRB1* alleles. Variants are named as described in <https://github.com/immunogenomics/HLA-TAPAS/wiki/HLA-imputation>.

Page 27

Frequency of combinations of amino acids at position 77 of the HLA-DRB1 protein and classical *HLA-DRB1* alleles. In the same format as page 26.

Reference list

1. Wingerchuk DM, Banwell B, Bennett JL, et al. International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. *Neurology* 2015; **85**(2): 177–89.
2. Morales J, Welter D, Bowler EH, et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biology* 2018; **19**(1): 21.
3. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 2012; **28**(24): 3326–8.
4. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010; **26**(22): 2867–73.
5. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**(1).
6. Conomos MP, Miller MB, Thornton TA. Robust Inference of Population Structure for Ancestry Prediction and Correction of Stratification in the Presence of Relatedness. *Genetic Epidemiology* 2015; **39**(4): 276–93.
7. Conomos Matthew P, Reiner Alexander P, Weir Bruce S, Thornton Timothy A. Model-free Estimation of Recent Genetic Relatedness. *The American Journal of Human Genetics* 2016; **98**(1): 127–48.
8. Gogarten SM, Sofer T, Chen H, et al. Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* 2019; **35**(24): 5346–8.
9. Chang CC. Data management and summary statistics with PLINK. *Methods Mol Biol* 2020; **2090**: 49–65.
10. Greer PJ, Sedlakova A, Ellison M, et al. A reassessment of Hardy-Weinberg equilibrium filtering in large sample Genomic studies. *medRxiv* 2024: 2024.02.07.24301951.
11. Grinde KE, Browning BL, Reiner AP, Thornton TA, Browning SR. Adjusting for principal components can induce spurious associations in genome-wide association studies in admixed populations. *bioRxiv* 2024: 2024.04.02.587682.
12. Kwong AM, Blackwell TW, LeFaive J, et al. Robust, flexible, and scalable tests for Hardy-Weinberg equilibrium across diverse ancestries. *Genetics* 2021; **218**(1).
13. 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**(18): 3426–40.e19.
14. Conomos Matthew P, Laurie Cecelia A, Stilp Adrienne M, et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *The American Journal of Human Genetics* 2016; **98**(1): 165–84.
15. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 2021; **590**(7845): 290–9.

16. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nature Genetics* 2016; **48**(10): 1284–7.
17. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics* 2014; **31**(5): 782–4.
18. Zheng X, Gogarten SM, Lawrence M, et al. SeqArray—a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics* 2017; **33**(15): 2251–7.
19. Aliferis CF, Statnikov A, Tsamardinos I, Mani S, Koutsoukos XD. Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: algorithms and empirical evaluation. *Journal of Machine Learning Research* 2010; **11**(1).
20. Steck H, Tresp V. Bayesian belief networks for data mining. Proceedings of the 2 workshop on data mining und data warehousing Als Grundlage Moderner Entscheidungsunterstützender Systeme; 1999: Citeseer; 1999. p. 145–54.
21. Ke X, Hunt S, Tapper W, et al. The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Human Molecular Genetics* 2004; **13**(6): 577–88.
22. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015; **31**(21): 3555–7.
23. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Frontiers in Genetics* 2020; **11**.
24. Estrada K, Whelan CW, Zhao F, et al. A whole-genome sequence study identifies genetic risk factors for neuromyelitis optica. *Nature Communications* 2018; **9**(1): 1929.
25. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190–1.
26. Ziyatdinov A, Kim J, Prokopenko D, et al. Estimating the effective sample size in association studies of quantitative traits. *G3 Genes|Genomes|Genetics* 2021; **11**(6).
27. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature genetics* 2012; **44**(4): 369–75.
28. Wang G, Sarkar A, Carbonetto P, Stephens M. A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 2020; **82**(5): 1273–300.
29. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the “Sum of Single Effects” model. *PLOS Genetics* 2022; **18**(7): e1010299.
30. Kerimov N, Hayhurst JD, Peikova K, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nature Genetics* 2021; **53**(9): 1290–9.
31. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLOS Genetics* 2021; **17**(9): e1009440.
32. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* 2015; **47**(3): 291–5.
33. Finucane HK, Reshef YA, Anttila V, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics* 2018; **50**(4): 621–9.
34. Consortium*† IMMSG, ANZgene, IIBDGC, WTCCC2. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 2019; **365**(6460): eaav7188.
35. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nature Genetics* 2015; **47**(11): 1236–41.
36. Beecham AH, Patsopoulos NA, Xifara DK, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature Genetics* 2013; **45**(11): 1353–60.