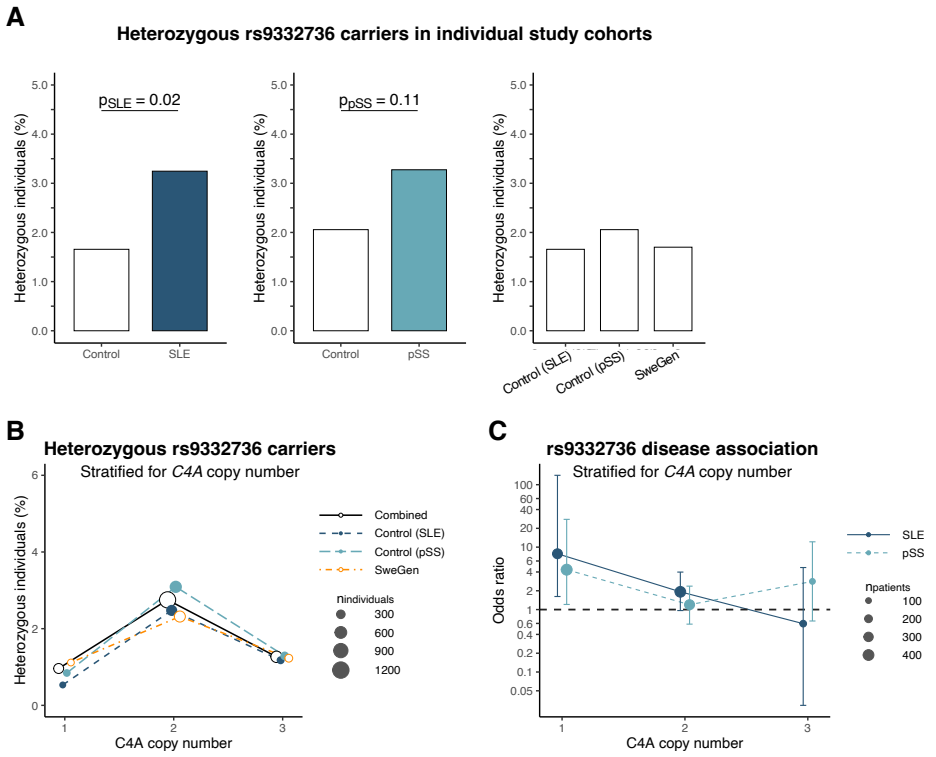
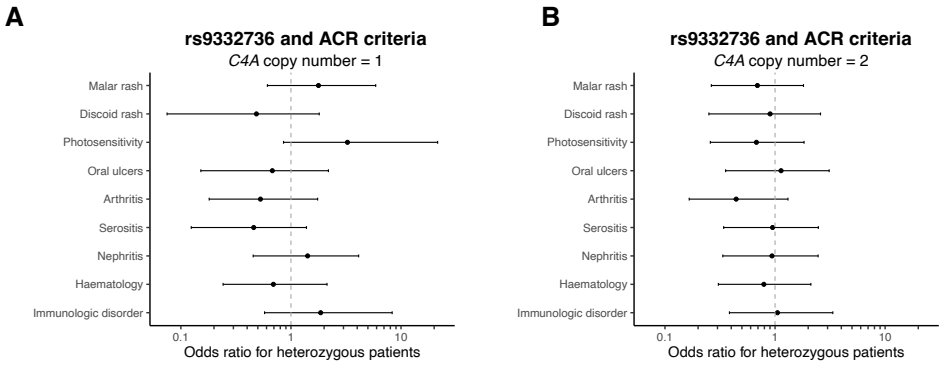
Supplementary Figure 1



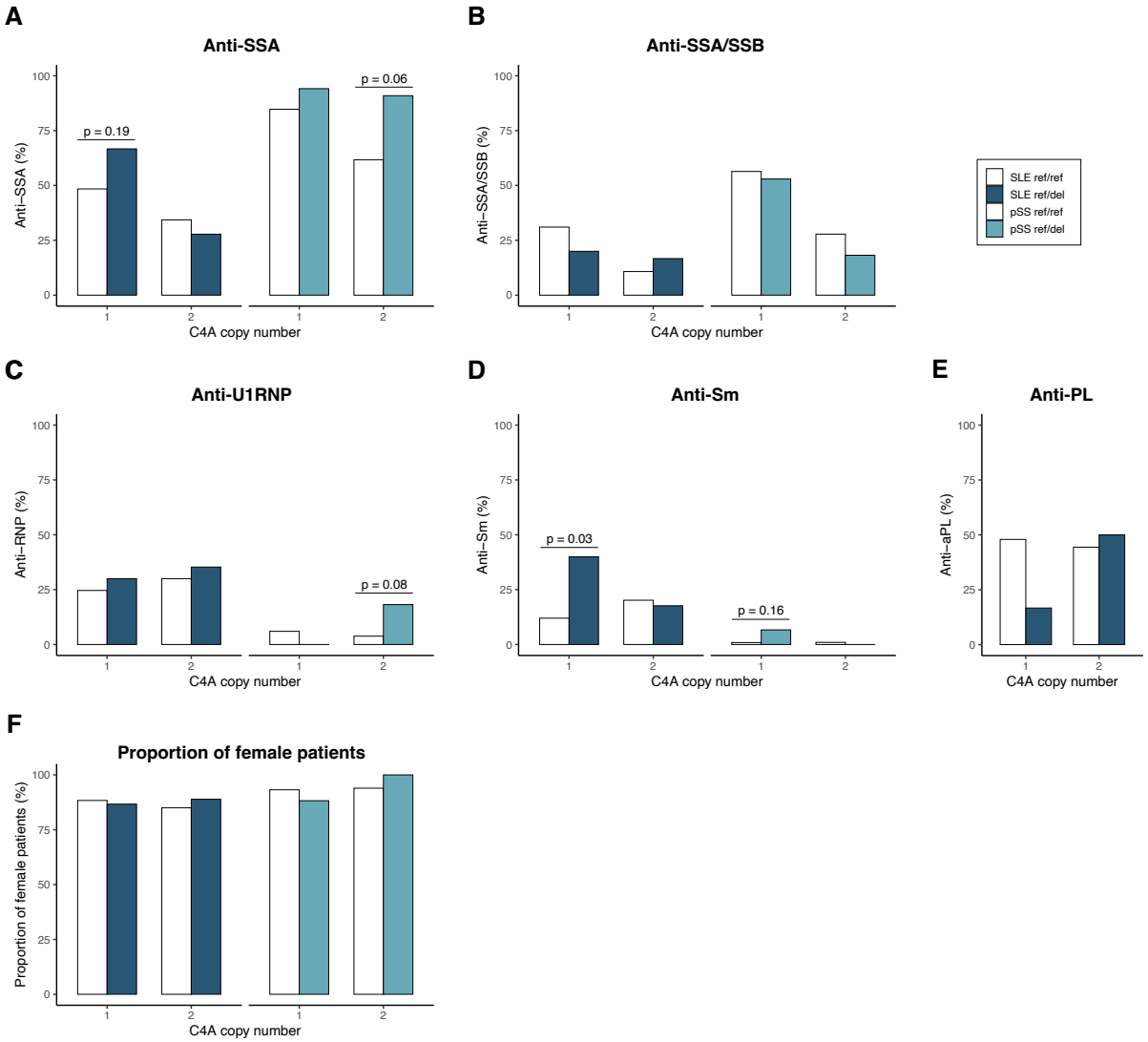
Supplementary Figure 1 Heterozygosity of the 28-bp *C2* deletion rs9332736 in individual cohorts
In the current study, we combined three cohorts of healthy/population controls from previous studies: from the SLE study (Sandling *et al.* 2021), from the pSS study (Thorlacius *et al.* 2021), and from SweGen (Ameur *et al.* 2017). The control groups for the SLE study and the pSS study were to a high extent shared between the two studies, meaning that 1,021 individuals were included in *both* studies, whereas 5 individuals were *only* included in the SLE study, and 243 individuals were *only* included in the pSS study. Here, we evaluate the prevalence of the rs9332736 variant in the individual cohorts.
A Prevalence of heterozygous carriers of the C2 loss-of-function variant rs9332736. Left graph: SLE (n = 955) and population controls (n = 1,026); middle graph: pSS (n = 916) and population controls (n = 1,264); right graph: SweGen (n = 1,000) and control cohorts from SLE and pSS studies, respectively. Analysis by logistic regression adjusting for sex. Individuals homozygous for the rs9332736 variant (n_{SLE} = 2; n_{pSS} = 1) have been excluded.
B Prevalence of heterozygous carriers of the C2 loss-of-function variant rs9332736 when stratifying for *C4A* copy number. The prevalence is shown for controls from the SLE study (n = 1,026), controls from pSS study (n = 1,264), SweGen (n = 1,000) and all controls combined (n = 2,262). Related individuals (n = 7) were excluded from the combined cohort, and note that and individuals and overlapping between the two control cohorts from SLE and pSS studies partially overlapped. The *C4A* copy number for individuals with the rs9332736 variant ranged between 1-3 copies.
C Association between rs9332736 and SLE/pSS compared to controls when stratifying for copy number of *C4A*. Analysed by logistic regression adjusting for sex and copy number of *C4B*. Size of point define number of patients at each copy number level.

Supplementary Figure 2



Supplementary Figure 2 Association between rs9332736 and clinical manifestations in SLE
Association to ACR criteria for patients heterozygous for the 28-bp *C2* deletion rs9332736 stratified for **(A)** *C4A* copy number of 1 and **(B)** *C4A* copy number of 2. Analysed by logistic regression, adjusting for sex and *C4B* copy number. Error bars represent 95% confidence interval. The two criteria 'neurological disorder' and 'antinuclear antibodies' are not analysed due to insufficient variation for patients heterozygous for the rs9332736 variant.

Supplementary Figure 3



Supplementary Figure 3 Association between rs9332736 and autoantibodies in SLE and pSS
Association to autoantibodies criteria for patients heterozygous for the 28-bp *C2* deletion rs9332736 (ref/ref vs. ref/del) stratified for *C4A* copy number.
A Anti-SSA/Ro autoantibodies ($n_{SLE,1\ C4A} = 360$, $n_{SLE,2\ C4A} = 455$, $n_{pSS,1\ C4A} = 429$, $n_{pSS,2\ C4A} = 374$).
B Anti-SSA/Ro and anti-SSB/La autoantibodies ($n_{SLE,1\ C4A} = 360$, $n_{SLE,2\ C4A} = 455$, $n_{pSS,1\ C4A} = 425$, $n_{pSS,2\ C4A} = 374$).
C Anti-RNP autoantibodies ($n_{SLE,1\ C4A} = 282$, $n_{SLE,2\ C4A} = 357$, $n_{pSS,1\ C4A} = 366$, $n_{pSS,2\ C4A} = 327$).
D Anti-Sm autoantibodies ($n_{SLE,1\ C4A} = 285$, $n_{SLE,2\ C4A} = 359$, $n_{pSS,1\ C4A} = 355$, $n_{pSS,2\ C4A} = 319$).
E Anti-phospholipid autoantibodies (aPL) in SLE patients ($n_{SLE,1\ C4A} = 175$, $n_{SLE,2\ C4A} = 199$).
F Proportion of female patients ($n_{SLE,1\ C4A} = 367$, $n_{SLE,2\ C4A} = 464$, $n_{pSS,1\ C4A} = 429$, $n_{pSS,2\ C4A} = 374$).
Unadjusted p-values (Fisher's exact test) are shown for $p < 0.20$.

Supplementary Figure 4

	Complete *C2* deficiency ($n_{patients}$ = 3)		
	pSS	**SLE**	**SLE**
Sex	Female	Female	Female
Age at diagnosis	40 years	52 years	30 years
Autoantibodies	(+) ANA, SSA, Scl-70 (low) (–) dsDNA, Sm, aPL, ANCA	(+) ANA, SSA, LAC (–) dsDNA, aCL, β2GPI, C1Q	(+) ANA, C1Q (–) dsDNA, aCL, β2GPI, RF
Complement	C2 deficient (< 6%) Normal C1q, C3, C4 Classical function: 0%	C2 deficient (< 6%) Normal C1q, C3, C4 Classical function: 10%	C2 deficient (<25%)# Normal C1q, C3, C4 Classical function: None
Kidney	No symptoms	Lupus nephritis eGFR 27	Normal eGFR 114
Other	Salivary gland biopsy: FS 4 Sicca, Raynaud, arthritis	Raynaud	Deceased (67 years) due to sepsis

Supplementary Figure 4 Clinical summary of patients with complete *C2* deficiency
Three patients homozygous for the 28-bp *C2* deletion rs9332736 were identified in the study. Selected clinical variables and results from serological analysis of complement are presented here for the three patients with complete *C2* deficiency. Reference interval for C2 concentration: 77-159%. Reference interval for classical complement function: 63-129%. #Complement analysis for patient has been performed in an earlier C2 assay with a higher detection threshold.
Abbreviations: SLE systemic lupus erythematosus, pSS primary Sjögren's syndrome, FS Greenspan focus score (lymphocytic infiltration in minor salivary gland biopsy), ANA antinuclear antibodies, Sm anti-Smith, aPL antiphospholipid antibodies, ANCA anti-neutrophil cytoplasmic antibodies, LAC lupus anticoagulant, aCL anti-cardiolipin antibodies, RF rheumatoid factor, eGFR estimated glomerular filtration rate.

Supplementary Table 1

Supplementary Table 1 Basic characteristics of study participants

Genetic analysis

	SLE	pSS	Control
n	958	911	2,262
Females	826 (86%)	849 (93%)	1,582 (70%)
Age at diagnosis	36 (3-85)	53 (14-90)[a]	-
Age at data abstraction	52 (18-94)	62 (19-92)	-
Age	-	-	54 (19-88)[b]

n (%) or mean (range) is shown

[a] Information missing for 1 individual

[b] Information missing for 1,178 individuals

Functional/clinical analysis

	SLE	pSS
n	1,088	973
Females	938 (86%)	908 (93%)
Age at diagnosis	36 (3-85)	53 (14-90)[a]
Age at data abstraction	52 (18-94)	61 (17-92)

n (%) or mean (range) is shown

[a] Information missing for 1 individual

Supplementary Table 2

Supplementary Table 2 rs9332736 allele frequency in 1000 Genomes Project

Super Population	Population	Description	rs9332736 ref/ref	ref/del	MAF	MAF
AFR African	ESN	Esan in Nigera	99	0	0	0
	GWD	Gambian in Western Division	113	0	0	
	LWK	Luhya in Webuye, Kenya	99	0	0	
	MSL	Mende in Sierra Leone	85	0	0	
	YRI	Yoruba in Ibadan, Nigera	108	0	0	
	ACB	African Carribean in Barbados	96	0	0	
	ASW	American's of African Ancestry in SW USA	61	0	0	
AMR Ad Mixed American	MXL	Mexican Ancestry from Los Angeles USA	62	2	0.0156	0.0043
	PUR	Puerto Rican from Puerto Rica	104	0	0	
	CLM	Colombian from Medellian, Colombia	93	1	0.0053	
	PEL	Peruvian from Lima, Peru	85	0	0	
EAS East Asian	CDX	Chinese Dai in Xishuanagbanna, China	93	0	0	0
	CHB	Han Chinese in Bejing, China	103	0	0	
	CHS	Southern Han Chinese	105	0	0	
	JPT	Japanese in Tokyo, Japan	104	0	0	
	KHV	Kinh in Ho Chi Minh City, Vietnam	99	0	0	
EUR European	CEU	Utah Residents with Northern/Western European ancestry	97	2	0.0101	0.0070
	IBS	Iberian population in Spain	106	1	0.0047	
	TSI	Toscani in Italia	106	1	0.0047	
	FIN	Finnish in Finland	97	2	0.0101	
	GBR	British in England and Scotland	90	1	0.0055	
SAS South Asian	PJL	Punjabi from Lahore, Pakistan	96	0	0	0
	BEB	Bengali from Bangladesh	86	0	0	
	GIH	Gujarati Indian from Houston, Texas	103	0	0	
	ITU	Indian Telugu from the UK	102	0	0	
	STU	Sri Lankan Tamil from the UK	102	0	0	

The 28-bp *C2* deletion rs9332736 was analysed in 2,504 unrelated high-coverage WGS samples from 1000 Genomes Project (Byrska-Bishop *et al.* 2021; bioRxiv) using GATK HaplotypeCaller. For comparison, the minor allele frequency (MAF) of Scandinavian controls in the current study was 0.0095 (43/2,262 heterozygous individuals).
Reference: Byrska-Bishop *et al.* High coverage whole genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *bioRxiv* 2021.02.06.430068. https://doi.org/10.1101/2021.02.06.430068

Supplementary Table 3

Supplementary Table 3 rs9332736 allele frequency in gnomAD

Population	ref/ref	ref/del	del/del	MAF
African/African American	12,455	25	0	0.001
East Asian	9,977	0	0	0
European	75,946	1,049	3	0.007
Latino/Admixed American	17,604	107	0	0.003
South Asian	15,305	3	0	0.0001

Frequency of the 28-bp *C2* deletion rs9332736 in the Genome Aggregation Database (gnomAD) (Karczewski *et al.* 2020). Information retrieved from https://gnomad.broadinstitute.org/ (variant: 6-31902065-ATGGTGGACAGGGTCAGGAATCAGGAGTC-A, GRCh37, v.2.1.1). For comparison, the minor allele frequency (MAF) of Scandinavian controls in the current study was 0.0095 (43/2,262 heterozygous individuals).

Reference: Karczewski *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443 (2020). https://doi.org/10.1038/s41586-020-2308-7

Supplementary Information

Combined genetic deficiencies of the classical complement pathway are strongly associated with both systemic lupus erythematosus and primary Sjögren's syndrome

Note: Part of the descriptions provided in the sections below are based on the Supplementary Information in Lundtoft *et al.* 2022 (1).

1. DNA sequencing, genotyping and quality control for DISSECT and SweGen

1.1. Cohorts

In the current study, we included Scandinavian SLE and pSS patients from the DISSECT study on systemic inflammatory autoimmune diseases together with healthy controls, as described previously (2, 3). The patients and controls had been analysed by targeted DNA sequencing. In addition, we included 1,000 population controls from the SweGen project that had been analysed by whole genome sequencing (WGS) (4). Sequencing and genotyping of the cohorts is described in the next sections.

For the genetic association analysis, we included all individuals that passed quality control in the different studies, but excluded related individuals (section 2). Quality control was performed both on variant- and individual-based level (i.e. population outliers were excluded; see next section). In addition to rs9332736 and *C4A* copy number results for the combined cohort, results for the individual cohorts are described in Supplementary Figure 1.

For functional and clinical analyses on the 28-bp *C2* deletion and *C4A* copy number, we included all patients with a quality-passed call for rs9332736 and *C4A* copy number in order to increase the power of the analyses.

1.2. DISSECT: Targeted DNA sequencing and genotyping

A custom SeqCap EZ Choice XL library (Roche NimbleGen) was designed to target exons and regulatory regions of 1,853 genes, as described in detail previously (5). Further, targeted sequencing of the samples included in the current study has been described previously (2, 3). In brief, 32 Mb were targeted for sequencing. Sequencing libraries were prepared by ultrasonification of DNA from whole blood to 400 bp fragments (Covaris E220) followed by barcoding (NEXTflex-96 DNA barcode adapters, Bio Scientific). Samples were pooled in batches of 8, hybridized (Roche NimbleGen) and sequenced with 100 bp paired-end reads using Illumina HiSeq 2500 version 3 or 4 chemistry.

Sequencing reads were mapped to the human hg19 reference using bwa mem (version 0.7.12), and duplicate reads were marked with Picard (version 1.92). Genotyping was performed using the GATK Best Practices workflow (GATK version 3.3.0) for variant discovery, indel realignment and base score recalibration prior to variant discovery using HaplotypeCaller in gVCF mode, excluding samples with a mean target coverage < 10x. Joint genotyping was performed separately for the SLE study (2) and the pSS study (3) using GATK GenotypeGVCFs, noting that healthy controls to a high extent overlapped between the two studies. Bi-allelic single-nucleotide variants were next passed on for recalibration of SNV quality scores using VariantRecalibrator with a filter at tranche level 99.0. Genotype calls with read depth < 8 and genotype Phred quality score < 20 were excluded using VCFtools.

The genetic structure of the study participants was analysed with LASER using the Human Genome Diversity Project (HGDP) as reference population (6, 7). Study participants > 5 standard deviations outside of the mean of the European sub-population of the HGDP reference set were excluded, followed by recursive exclusion of subjects exceeding > 5 standard deviations of the remaining study subjects. Duplicate and first-degree related individuals were excluded based on relatedness analysed using KING (8). An extra filter on rate of missing data, heterozygosity ratio, transition-transversion ratio and singleton counts was applied to exclude extreme sample outliers (2). Finally, samples with a call rate < 80% were removed.

1.3. SweGen: Whole genome sequencing and genotyping

DNA sequencing and genotyping of 1,000 individuals from Sweden has been described in detail previously (4). Briefly, DNA was fragmented into 350 bp insert sizes, and paired-end sequencing with 150 bp read length was performed on Illumina HiSeq X with v2.5 sequencing chemistry.

Sequencing reads were mapped to GRCh37 using bwa mem and subsequently genotyped according to GATK Best Practices workflow using GATK version 3.3, including indel realignment, mark of duplicate reads (Picard), and base quality score recalibration. GATK HaplotypeCaller was used to genotype individual samples in gVCF mode, followed by joint genotyping using CombineGVCFs and GenotypeGVCFs. Finally, SNVs and indels were recalibrated using GATK VQSR.

1.4. Exclusion of related individuals from the combined cohort

From the combined cohort of patients and controls, we excluded first-degree related individuals using KING version 2.2.6 (8). SNVs overlapping between all studies were used as input for the analysis, and one individual from each pair of related individuals were sequentially removed first from the SweGen cohort (n = 4), followed by DISSECT controls (n = 3) and finally pSS patients (n = 6). None of the related pairs carried the 28-bp *C2* deletion rs9332736.

2. Genotyping of the 28-bp *C2* deletion rs9332736

As the previous analysis of genetic variation in the DISSECT project with targeted sequencing data comprised SNVs only (2, 3), a focused re-analysis was performed in order to genotype the 28-bp deletion rs9332736 (GRCh37.p13 chromosome 6 NC_000006.11:g.31902068_31902095del) in *C2*. Using GATK HaplotypeCaller (version 4.1.8.1), we analysed genetic variation in the entire *C2* gene ± 1000 bp (hg19; chr6:31864560-31914449). Individual calls for rs9332736 were merged and genotyped using GATK CombineGVCFs and GenotypeGVCF, and subsequently filtered (read depth ≥ 8, genotyping quality ≥ 20) using bcftools (version 1.12).

For the SweGen WGS data, SNVs and indels had been genotyped using GATK HaplotypeCaller followed by CombineGVCFs, GenotypeGVCF and variant quality score recalibration (VQSR) as described previously (4).

Plots showing read depth and fraction of reference/alternative allele of the 28-bp deletion rs9332736 from DISSECT targeted sequencing calls and SweGen WGS calls are shown in Fig. 1 and Fig. 2, respectively. One SLE patient with a heterozygous call for rs9332736 that clustered with homozygous calls was excluded from the analysis as a conservative measure (Fig. 1). The patient had two copies of *C4A*.

The *C2* variant rs9332736 did not deviate from Hardy-Weinberg equilibrium for population controls (p = 0.65).

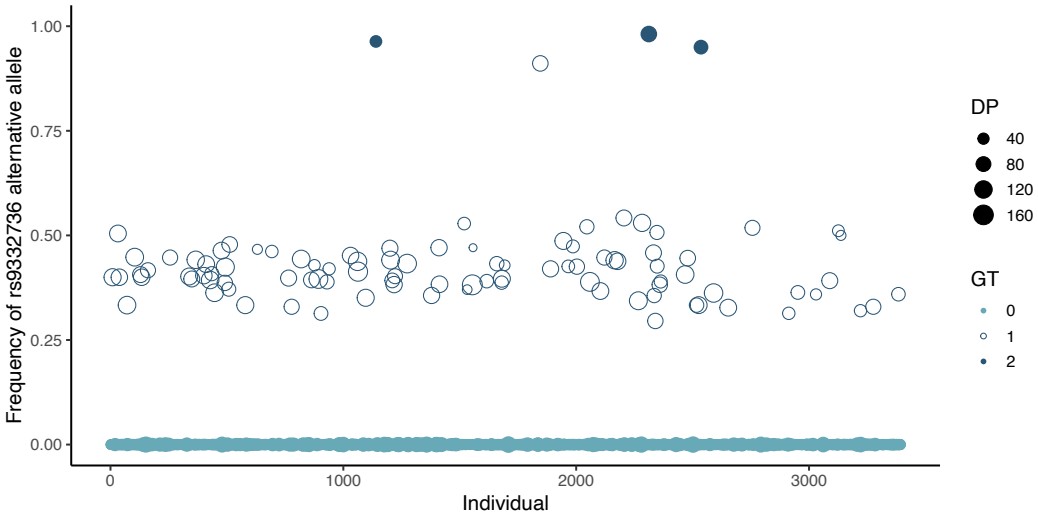


Fig. 1 Read depth (DP) and frequency of the 28-bp *C2* deletion rs9332736 in reads from DISSECT targeted sequencing data (n = 3,393). The genotype (GT) call is indicated in the plot. One SLE patient that clustered with the homozygous rs9332736 carriers but with a heterozygous call was excluded from the analysis (see text).

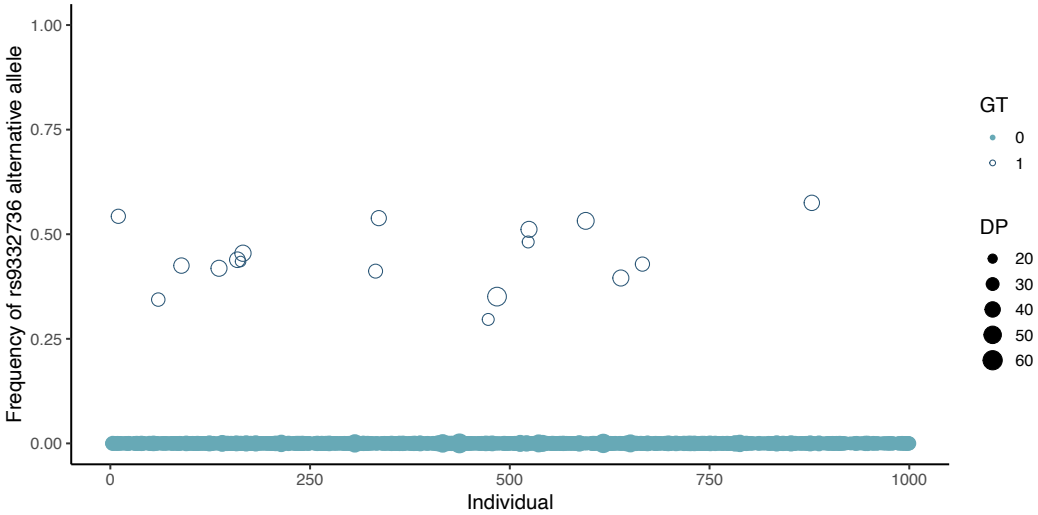


Fig. 2 Read depth (DP) and frequency of the 28-bp *C2* deletion rs9332736 in reads from SweGen WGS data (n = 1,000). The genotype call (GT; 0-2 rs9332734 alleles) is indicated in the plot.

3. Analysis of *C4* copy number

Analysis of *C4* copy number from both targeted sequencing data and WGS data has been described and validated in detail previously (1), and a brief description is provided in this section.

3.1. Structure of C4A/C4B

The human paralogous *C4* genes, *C4A* and *C4B*, are located in the *HLA* class III region on the p arm of chromosome 6, centromeric to *HLA* class I and telomeric to *HLA* class II. The two *C4* genes are both 20.6 kb long and code for 41 exons (Fig. 3). The reference sequences of the two genes differ at 18 positions (Table 1), thereby being 99.91% identical. Five nucleotide variants – leading to 4 amino acid substitutions in exon 26 (PCPVLD vs. LSPVIH) – are used to distinguish *C4A* and *C4B* (Table 1). Some *C4* genes may contain a ~6 kb human endogenous retroviral (*HERV*) insertion between exon 9 and 10 (Fig. 3), but considering that this region

has not been targeted for sequencing as part of this study, copy number variation of the *HERV* insertion is not part of the current *C4* analysis.

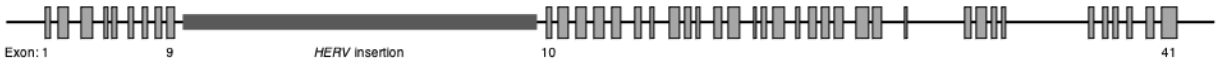


Fig. 3 Structure of the paralogous *C4* genes, *C4A* and *C4B*. The *C4* genes may contain a ~6 kb human endogenous retroviral (*HERV*) insertion.

Table 1 Variants differing between the reference sequence for *C4A* and *C4B*. The 5 nucleotide variants in exon 26 (causing 4 amino acid substitutions) used to define *C4A* and *C4B*, respectively, are marked in bold.

Position (GRCh37)		Position (GRCh38)		Allele		
C4A	*C4B*	*C4A*	*C4B*	*C4A*	*C4B*	Exon/Intron
31962174	31994912	31994397	32027135	A	G	Intron 20
31962401	31995139	31994624	32027362	G	A	Ala/Thr (exon 21)
31963559	31996297	31995782	32028520	A	G	Asp/Gly (exon 25)
31963860	**31996598**	**31996083**	**32028821**	**C**	**T**	**Pro/Leu (exon 26)**
31963863	**31996601**	**31996086**	**32028824**	**G**	**C**	**Cys/Ser (exon 26)**
31963871	**31996609**	**31996094**	**32028832**	**T**	**A**	**Leu/Ile (exon 26)**
31963874	**31996612**	**31996097**	**32028835**	**G**	**C**	**Asp/His (exon 26)**
31963876	**31996614**	**31996099**	**32028837**	**C**	**T**	
31964228	31996966	31996451	32029189	A	G	Asn/Ser (exon 28)
31964316	31997054	31996539	32029277	G	C	Ala/Ala (exon 28)
31964321	31997059	31996544	32029282	T	C	Val/Ala (exon 28)
31964330	31997068	31996553	32029291	T	G	Leu/Arg (exon 28)
31964331	31997069	31996554	32029292	C	G	
31964391	31997129	31996614	32029352	C	G	Intron 28
31964394	31997132	31996617	32029355	TC	T	Intron 28
31964785	31997522	31997008	32029745	T	G	Ser/Ala (exon 29)
31965242	31997979	31997465	32030202	T	C	Intron 30
31965383	31998120	31997606	32030343	A	G	Intron 30

3.2. Calling C4 copy number

C4 copy number was estimated using GATK GermlineCNVCaller (version 4.1.8.1), which is a read depth-based method for analysis of copy number variation in WES/targeted sequencing data using bam files as input. Prior to analysis, reads mapped to the *C4A*/*C4B* regions ± 500bp (hg19, chr6:31949334-32003695) were extracted (samtools version 1.10) and remapped (bwa mem version 0.7.17) to the reference sequence for chromosome 6 in which *C4A* ± 1,000 bp (chr6:31948834-31971457) had been masked. Next, the *C4* reads mapped to the *C4A*-masked reference were merged with chromosome 6 reads outside the *C4A*/*C4B* region ± 1,000 bp (chr6:1-31948834 and chr6:32004195-171115067). Before analysis in the GermlineCNVCaller pipeline, duplicate reads were marked using Picard (version 2.20.4).

Samples were analysed using the GATK GermlineCNVCaller pipeline in cohort mode with batches of size ~300. Forty samples with known *C4* copy number were included in all batches to allow for quality control and normalisation (see below). Intervals on chromosome 6 targeted for sequencing were first split to have a maximum size of 5,000 bp. Intervals in the *C4B* region were manually defined to cover the relevant regions (chr6:31982572-31984923, chr6:31991707-31994992, chr6:31994993-31998278, chr6:31999328-32000075, chr6:32001567-32003195), and a total number of 5,478 intervals on chromosome 6 were prepared using GATK

PreprocessIntervals using default settings for targeted sequencing data. In the next step, the number of reads was analysed sample-wise for all intervals using CollectReadCounts, followed by AnnotateIntervals, FilterIntervals [--extreme-count-filter-maximum-percentile 100], DetermineGermlineContigPloidy, GermlineCNVCaller [--max-copy-number 8], and PostprocessGermlineCNVCalls (alternative settings defined in brackets).

The output from GermlineCNVCaller is a 'denoised copy ratio' for each interval across all individual samples. The total copy number of *C4* was estimated for each sample based on the average denoised copy ratio of the 5 *C4B* intervals. The copy number estimate was next normalised within each batch by linear regression using the samples with known *C4* copy number. Combining *C4* copy number estimates from all samples showed a multimodal distribution (Fig. 4), and the continuous estimate was rounded to the nearest integer copy number value.

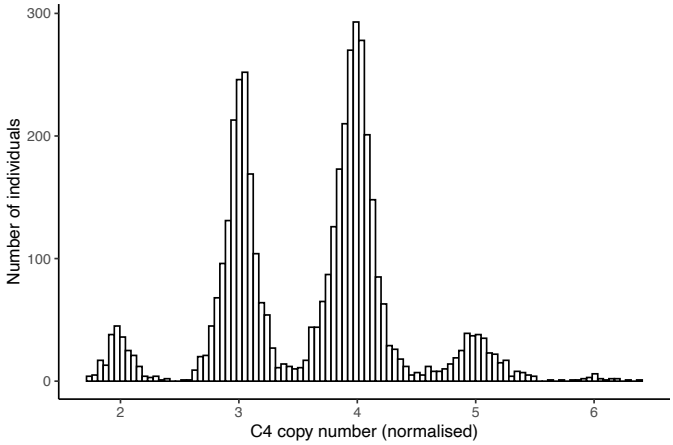


Fig. 4 *C4* copy number estimates of for healthy controls and patients with SLE and pSS (n = 4,389). Copy number calls from DISSECT and SweGen have been combined. Three individuals with copy number ≥ 7 are not included in the plot.

The proportion of *C4A* and *C4B* genes among the total number of *C4* genes was estimated based on the average read depth of the 5 paralog-specific variants (*C4B*: chr6 position 31996598T/C, 31996601C/G, 31996609A/T, 31996612C/G and 31996614T/C) analysed using GATK HaplotypeCaller. By plotting the estimate for total *C4* copy number against the read depth of *C4A*-specific variants relative to the total read depth of both *C4A*- and *C4B*-specific variants, samples generally clustered on the integer combinations of *C4A*/*C4B* copy number possible for each *C4* copy number level (Fig. 5).

Based on the total *C4* copy number, the integer copy number of *C4A* and *C4B* were calculated from their relative *C4A*-specific read depth using the relation: $C4 = C4A + C4B$. *C4A*/*C4B* copy number was not estimated for samples with a total read depth < 10 of the *C4A*-/*C4B*-defining variants, meaning that *C4A*/*C4B* copy number was not estimated for 28 individuals.

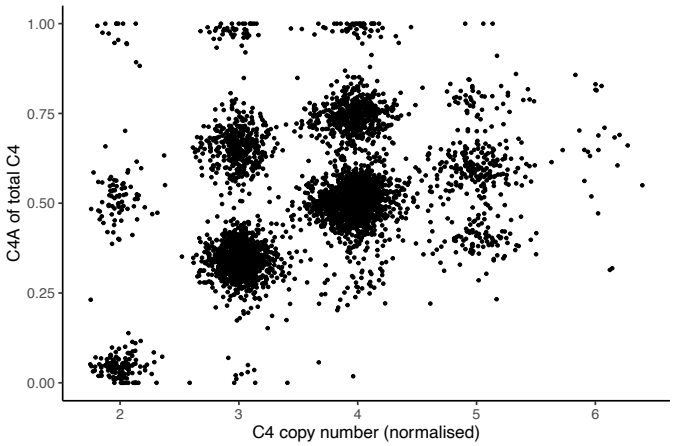


Fig. 5 *C4* copy number estimates plotted against the proportion of the read depth of *C4A*-specific variants relative to the total read depth of C4A-/C4B-specific variants (n = 4,361). Copy number calls from DISSECT and SweGen have been combined. Three individuals with copy number ≥ 7 are not included in the plot, and 28 individuals with low read depth of *C4A*/*C4B*-defining nucleotides have been excluded.

For SweGen WGS data, reads mapped to a 5 Mb region of the *HLA* region (hg19; chr6:29000000-34000000), excluding duplicate reads, were extracted and remapped to the reference for chromosome 6, in which *C4A* ± 1,000 bp had been masked. Intervals of 1000 bp size were generated for the 5 Mb HLA region using PreprocessIntervals, and intervals in *C4B* region were manually defined to ~1,000 bp intervals in the covering chr6:31982572-31984923 (*C4B* exon 1-9, 3 intervals), chr6:31991707-32003195 (*C4B* exon 10-41, 12 intervals), and chr6:31985199–31991567 (*HERV* sequence (9), 7 intervals). The residual analysis in GermlineCNVCaller was done as described above for targeted sequencing data. *C4* copy number was estimated based on the average denoised copy ratio of the 15 *C4B* intervals.

4. Calling of *HLA*

Analysis of *HLA* from DNA sequencing data has been described and validated in detail previously (1). Briefly, *HLA* alleles of the 6 genes *HLA-A*, *-B*, *-C*, *-DPB1*, *-DQB1* and *-DRB1* were called at 2-field (i.e. 4-digit) resolution from sequencing data using xHLA (10). Prior to analysis, reads in the extended *HLA* region (chr6:29-34mb) and unmapped reads were remapped to chromosome 6 of the GRCh38 reference, and duplicate reads were discarded.

5. Consortia

5.1. The DISSECT consortium

Lars Rönnblom (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Gunnel Nordmark (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Ingrid E. Lundberg (Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden), Johanna K. Sandling (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Pascal Pucholt (Department of Medical Sciences, Rheumatology, Uppsala University, Sweden), Sergey V. Kozyrev (Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden), Maija-Leena Eloranta (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Matteo Bianchi (Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University,

Uppsala, Sweden), Roland Jonsson (Broegelmann Research Laboratory, Department of Clinical Science, University of Bergen, Bergen, Norway), Roald Omdal (Department of Rheumatology, Stavanger University Hospital, Stavanger, Norway and Broegelmann Research Laboratory, Department of Clinical Science, University of Bergen, Bergen, Norway), Ann-Christine Syvänen (Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden), Andreas Jönsen (Department of Clinical Sciences Lund, Rheumatology, Lund University, Skåne University Hospital, Lund, Sweden), Iva Gunnarsson (Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden), Elisabet Svenungsson (Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden), Solbritt Rantapää-Dahlqvist (Department of Public Health and Clinical Medicine/Rheumatology, Umeå University, Umeå, Sweden), Anders A. Bengtsson (Department of Clinical Sciences Lund, Rheumatology, Lund University, Skåne University Hospital, Lund, Sweden), Christopher Sjöwall (Department of Biomedical and Clinical Sciences, Division of Inflammation and Infection, Linköping University, Linköping, Sweden), Dag Leonard (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Kerstin Lindblad-Toh (Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden and Broad Institute of MIT and Harvard, Cambridge, MA, USA), Jennifer R. S. Meadows (Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden), Jessika Nordin (Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden), Marie Wahren-Herlenius (Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden and Broegelmann Research Laboratory, Department of Clinical Science, University of Bergen, Norway), Sule Yavuz (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Daniel Hammenfors (Department of Rheumatology, Haukeland University Hospital, Bergen, Norway), Elke Theander (Department of Rheumatology, Skåne University Hospital Malmö/Lund University, Lund, Sweden), Eva Baecklund (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Guðný Ella Thorlacius (Department of Medicine, Unit for Experimental Rheumatology, Karolinska Institutet, Stockholm, Sweden), Helena Enocsson (Department of Biomedical and Clinical Sciences, Division of Inflammation and Infection, Linköping University, Linköping, Sweden), Helena Forsblad-d'Elia (Department of Rheumatology and Inflammation Research, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden), Johan G. Brun (Department of Rheumatology, Haukeland University Hospital, University of Bergen, Bergen, Norway), Juliana Imgenberg-Kreuz (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Karl A. Brokstad (Broegelmann Research Laboratory, Department of Clinical Science, University of Bergen, Bergen, Norway), Kathrine Skarstein (The Gade Laboratory for Pathology, Department of Clinical Medicine, University of Bergen, Norway), Katrine Brække Norheim (Department of Rheumatology, Stavanger University Hospital, Stavanger, Norway and Institute of Clinical Science, University of Bergen, Bergen, Norway), Lilian Vasaitis (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Malin V. Jonsson (Section for Oral and Maxillofacial Radiology, Department of Clinical Dentistry, University of Bergen, Bergen, Norway), Marika Kvarnström (Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden and Academic Specialist Center, Center for Rheumatology, Stockholm Health Services, Region Stockholm, Stockholm, Sweden), Per Eriksson (Department of Biomedical and Clinical Sciences, Division of Inflammation and Infection, Linköping University, Linköping, Sweden), Sara Bucher (Department of Rheumatology, Faculty

of Medicine and Health, Örebro University, Örebro, Sweden), Silke Appel (Broegelmann Research Laboratory, Department of Clinical Science, University of Bergen, Bergen, Norway), Svein Joar Johnsen (Department of Rheumatology, Stavanger University Hospital, Stavanger, Norway), Thomas Mandl (Department of Clinical Sciences Malmö, Division of Rheumatology, Lund University, Malmö, Sweden), Lara Adnan Aqrawi (Department of Oral Surgery and Oral Medicine, Institute of Clinical Odontology, University of Oslo, Oslo, Norway and Department of Health Sciences, Kristiania University College, Oslo, Norway), Janicke Liaaen Jensen (Department of Oral Surgery and Oral Medicine, Institute of Clinical Odontology, University of Oslo, Oslo, Norway), Øyvind Palm (Department of Rheumatology, Oslo University Hospital, Oslo, Norway), Fabiana H.G. Farias (Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden, and Department of Psychiatry, Washington University, St. Louis, MO, USA), Leonid Padyukov (Division of Rheumatology, Department of Medicine, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden), Johanna Dahlqvist (Science for Life Laboratory, Department of Medical Sciences and Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden), Daniel Eriksson (Department of Medicine (Solna), Karolinska Institutet, and Department of Endocrinology, Metabolism and Diabetes Karolinska University Hospital, Stockholm, Sweden), Argyri Mathioudaki (Department of Medical Sciences, Array & Analysis Facility, Uppsala University, Uppsala, Sweden), Albin Björk (Department of Medicine, Unit for Experimental Rheumatology, Karolinska Institutet, Stockholm, Sweden).

5.2. The ImmunoArray development consortium

Kerstin Lindblad-Toh (Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden and Broad Institute of MIT and Harvard, Cambridge, MA, USA), Gerli Rosengren Pielberg (Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden), Anna Lobell (Office for Medicine and Pharmacy, Uppsala University, Uppsala, Sweden), Åsa Karlsson (Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden), Göran Andersson (Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden), Kerstin M. Ahlgren (Department of Surgical Sciences, Uppsala University, Uppsala, Sweden), Lars Rönnblom (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Maija-Leena Eloranta (Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden), Nils Landegren (Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden and Centre for Molecular Medicine, Department of Medicine (Solna), Karolinska Institutet, Stockholm, Sweden), Olle Kämpe (Department of Medicine (Solna), Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden, Department of Endocrinology, Metabolism and Diabetes Karolinska University Hospital, Stockholm, Sweden and KG Jebsen Center for autoimmune diseases, University of Bergen, Norway), Peter Söderkvist (Division of Cell Biology, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden).

6. References

1. Lundtoft C, Pucholt P, Martin M, Bianchi M, Lundström E, Eloranta M-L, et al. Complement *C4* copy number variation is linked to SSA/Ro and SSB/La autoantibodies in systemic inflammatory autoimmune diseases. Arthritis & Rheumatology. 2022.
2. Sandling JK, Pucholt P, Hultin Rosenberg L, Farias FHG, Kozyrev SV, Eloranta M-L, et al. Molecular pathways in patients with systemic lupus erythematosus revealed by gene-centred DNA sequencing. Annals of the Rheumatic Diseases. 2021;80(1):109-17.

3. Thorlacius GE, Hultin-Rosenberg L, Sandling JK, Bianchi M, Imgenberg-Kreuz J, Pucholt P, et al. Genetic and clinical basis for two distinct subtypes of primary Sjögren's syndrome. Rheumatology. 2021;60(2):837-48.
4. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. European Journal Of Human Genetics. 2017;25:1253.
5. Eriksson D, Bianchi M, Landegren N, Nordin J, Dalin F, Mathioudaki A, et al. Extended exome sequencing identifies BACH2 as a novel major risk locus for Addison's disease. Journal of Internal Medicine. 2016;280(6):595-608.
6. Wang C, Zhan X, Bragg-Gresham J, Kang HM, Stambolian D, Chew EY, et al. Ancestry estimation and control of population stratification for sequence-based association studies. Nature Genetics. 2014;46(4):409-15.
7. Wang C, Zhan X, Liang L, Abecasis Gonçalo R, Lin X. Improved Ancestry Estimation for both Genotyping and Sequencing Data using Projection Procrustes Analysis and Genotype Imputation. The American Journal of Human Genetics. 2015;96(6):926-37.
8. 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.
9. Kamitaki N, Sekar A, Handsaker RE, de Rivera H, Tooley K, Morris DL, et al. Complement genes contribute sex-biased vulnerability in diverse disorders. Nature. 2020;582(7813):577-81.
10. Xie C, Yeo ZX, Wong M, Piper J, Long T, Kirkness EF, et al. Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. Proceedings of the National Academy of Sciences. 2017;114(30):8059.

Supplemental data for Figure 1

A

disease	n heterzygotes	n total
Control	43	2262
SLE	31	955
pSS	30	910

B

disease	*C4A* copy number	OR (95% CI)	p-value	n patients	n controls
SLE	1	4.40 (1.58-15.73)	0.01	333	416
pSS	1	3.83 (1.31-14.12)	0.02	408	416
SLE	2	1.71 (0.95-3.03)	0.07	394	1163
pSS	2	1.33 (0.67-2.52)	0.39	342	1163
SLE	3	0.68 (0.04-4.03)	0.72	142	553
pSS	3	3.40 (0.84-12.45)	0.07	102	553

C

disease	rs9332736	*C4A* copy number	OR (95% CI)	n patients	n controls
SLE	0	0	7.49 (4.80-11.94)	73	31
SLE	1	0	NA		
pSS	0	0	4.93 (3.03-8.15)	46	31
pSS	1	0	NA		
SLE	0	1	2.37 (1.96-2.88)	333	416
SLE	1	1	10.22 (3.52-37.03)		
pSS	0	1	3.41 (2.81-4.15)	408	416
pSS	1	1	13.05 (4.47-48.39)		
SLE	0	2	Reference	394	1163
SLE	1	2	1.58 (0.84-2.88)		
pSS	0	2	Reference	342	1163
pSS	1	2	1.14 (0.54-2.28)		

D

Chromosome	Position (bp)	R^2
6	31486405	0.893
6	31704411	0.893
6	32370624	0.789
6	31418281	0.735
6	31340001	0.735
6	31123434	0.674
6	31242329	0.674
6	31175118	0.674
6	30998558	0.600
6	30996325	0.600
6	32778203	0.530

Supplemental data for Figure 1: Heterozygosity of the 28-bp *C2* deletion rs9332736 in SLE and pSS

A Prevalence of heterozygous carriers of the *C2* loss-of-function variant rs9332736 in SLE patients, pSS patients and controls. Comparison between patients and controls performed by logistic regression adjusting for sex. Individuals homozygous for the rs9332736 variant ($n_{SLE} = 2$; $n_{pSS} = 1$) have been excluded.

B Association between rs9332736 and SLE/pSS compared to controls when stratifying for copy number of *C4A*.

C Combined effect of *C4A* copy number and rs9332736 heterozygosity in relation to a *C4A* copy number of 2 and normal *C2*. Due to rs9332736 being segregated with *C4A*, no individuals heterozygous for rs9332736 have a *C4A* copy number of 0.

(B and C) Analysed by logistic regression adjusting for sex and copy number of *C4B*. Total numbers of patients/controls are indicated at each level.

D Linkage disequilibrium (LD; R^2) between the 28-bp *C2* deletion rs9332736 and *HLA* alleles/bialleic SNPs in the *HLA* region. The 10 variants with strongest LD are listed (position refers to hg19). SweGen WGS samples (n = 1,000) were used for the LD estimation.

Supplemental data for Supplementary Figure 1

A

disease	n heterzygotes	n total
Control (SLE)	17	1026
SLE	31	955
Control (pSS)	26	1264
pSS	30	916
Control (SLE)	17	1026
Control (pSS)	26	1264
SweGen	17	1000

B

C4A copy number	disease	n heterozygotes	n total
0	Combined	0	31
0	Control (SLE)	0	14
0	Control (pSS)	0	18
0	SweGen	0	13
1	Combined	4	416
1	Control (SLE)	1	188
1	Control (pSS)	2	237
1	SweGen	2	179
2	Combined	32	1163
2	Control (SLE)	13	525
2	Control (pSS)	20	647
2	SweGen	12	518
3	Combined	7	553
3	Control (SLE)	3	255
3	Control (pSS)	4	310
3	SweGen	3	244
4	Combined	0	92
4	Control (SLE)	0	42
4	Control (pSS)	0	50
4	SweGen	0	41
5	Combined	0	7
5	Control (SLE)	0	2
5	Control (pSS)	0	2
5	SweGen	0	5

C

disease	*C4A* copy number	OR (95% CI)	n patients	n controls
SLE	1	7.83 (1.62-141.35)	333	188
pSS	1	4.35 (1.21-27.85)	412	237
SLE	2	1.92 (0.96-3.99)	394	525
pSS	2	1.20 (0.58-2.37)	343	647
SLE	3	0.60 (0.03-4.71)	142	255
pSS	3	2.83 (0.66-12.17)	102	310

Supplemental data for Supplementary Figure 1: Heterozygosity of the 28-bp *C2* deletion rs9332736 in individual cohorts
A Prevalence of heterozygous carriers of the *C2* loss-of-function variant rs9332736 in original study cohorts. Individuals homozygous for the rs9332736 variant (n_{SLE} = 2; n_{pSS} = 1) have been excluded.
B Prevalence of heterozygous carriers of the *C2* loss-of-function variant rs9332736 when stratifying for *C4A* copy number. The prevalence is shown for controls from the SLE study (n = 1,026), controls from pSS study (n = 1,264), SweGen (n = 1,000) and all controls combined (n = 2,262). Related individuals (n = 7) were excluded in the combined cohort, and note that and individuals and overlapping between the two control cohorts from SLE and pSS studies partially overlapped.
C Association between rs9332736 and SLE/pSS compared to controls when stratifying for copy number of *C4A*. Analysed by logistic regression adjusting for sex and copy number of *C4B*.

Supplemental data for Supplementary Figure 2

A

	ACR criteria	OR (95% CI)	n
C4A copy number = 1	Malar rash	1.78 (0.61-5.89)	367
	Discoid rash	0.49 (0.07-1.81)	367
	Photosensitivity	3.26 (0.86-21.51)	367
	Oral ulcers	0.68 (0.15-2.19)	367
	Arthritis	0.53 (0.18-1.75)	367
	Serositis	0.46 (0.12-1.38)	367
	Neptritis	1.42 (0.45-4.12)	367
	Haematology	0.69 (0.24-2.13)	367
	Immunologic disorder	1.86 (0.58-8.31)	367

	ACR criteria	OR (95% CI)	n
C4A copy number = 2	Malar rash	0.69 (0.26-1.81)	464
	Discoid rash	0.90 (0.25-2.59)	464
	Photosensitivity	0.68 (0.26-1.83)	464
	Oral ulcers	1.13 (0.36-3.10)	464
	Arthritis	0.44 (0.17-1.31)	464
	Serositis	0.95 (0.34-2.47)	464
	Neptritis	0.94 (0.34-2.46)	464
	Haematology	0.79 (0.31-2.12)	464
	Immunologic disorder	1.05 (0.39-3.35)	464

Supplemental data for Supplementary Figure 2: Association between rs9332736 and clinical manifestations in SLE
Association to ACR criteria for patients heterozygous for the 28-bp *C2* deletion rs9332736 stratified for **(A)** *C4A* copy number of 1 and **(B)** *C4A* copy number of 2. Analysed by logistic regression, adjusting for sex and *C4B* copy number. The two criteria 'neurological disorder' and 'antinuclear antibodies' are not analysed due to insufficient variation for patients heterozygous for the rs9332736 variant.