

## Supplementary files

## Supplementary table 1. Study design

Population	GWAS genotyping platform	N Case/Controls After genotyping quality controls	Genotyped SNPs	Analyzed SNPs (overlapping between populations)	Statistical Power*
USA	Illumina Human 550K /(1) breast cancer controls CGEMS, (2) prostate cancer controls CGEMS; and (3) Illumina controls (ref)	1486/3477	488.793	279.621	0.99
Spain	Illumina Human CNV370K BeadChip/ Illumina Human CNV370K BeadChip	364/384	322.967	279.621	0.90
Germany	Illumina Human CNV370K BeadChip/ Illumina Human550K (ref)	270/671	308.299	279.621	0.92
The Netherlands	Illumina Human CNV370K BeadChip/ Illumina Human550K (ref)	176/639	308.349	279.621	0.81
Total	-	2296/5171	-		0.99

\*for detecting and OR of 1.5 with a MAF of 0.20

**Supplementary table 2. SNPs associated with SSc at genome-wide significance within the MHC region.**

CHR	SNP	Closest Gene	BP	Minor Allele	MAF	GC	PC	OR	CI 95%
						Corrected P Value	Corrected P value		
6	rs6457617	HLA-DQB1	32,771,829	C	0.466	3.98E-17	3.85E-19	0.729	0.68-0.78
6	rs2856705	HLA-DQB1	32,778,934	A	0.096	2.19E-14	3.71E-14	0.591	0.52-0.67
6	rs9275390	HLA-DQB1	32,777,134	C	0.271	2.17E-11	1.87E-10	1.314	1.22-1.42
6	rs5000634	HLA-DQB1	32,771,542	C	0.403	3.88E-11	9.13E-13	1.283	1.20-1.38
6	rs443198	NOTCH4	32,298,384	C	0.355	5.98E-10	6.82E-12	0.784	0.73-0.84
6	rs2071286	NOTCH4	32,287,874	A	0.25	2.68E-09	1.67E-10	1.284	1.19-1.39
6	rs2516399	MICB	31,589,278	C	0.105	5.68E-09	6.27E-09	0.688	0.61-0.78
6	rs1521	MICA	31,458,683	C	0.235	7.18E-09	8.96E-09	0.769	0.71-0.84
6	rs9275312	HLA-DQB1	32,773,706	G	0.141	8.04E-09	3.70E-08	1.345	1.22-1.48
6	rs479536	NOTCH4	32,301,656	T	0.058	1.59E-08	1.50E-08	0.615	0.52-0.72
6	rs2071295	TNXB	32,146,678	A	0.34	2.22E-08	2.71E-09	1.242	1.15-1.34
6	rs2523477	MICA	31,468,368	G	0.082	2.64E-08	1.94E-07	0.669	0.58-0.77
6	rs12153855	TNXB	32,182,782	C	0.098	2.70E-08	5.08E-08	0.692	0.61-0.78
6	rs2844494	MICB	31,591,394	G	0.312	2.99E-08	3.22E-08	0.798	0.74-0.86
6	rs2239689	TNXB	32,138,262	T	0.342	4.01E-08	4.64E-09	1.237	1.15-1.33
6	rs3129871	HLA-DRA	32,514,320	A	0.343	4.92E-08	4.95E-08	0.806	0.75-0.87
6	rs1035798	AGER	32,259,200	T	0.278	5.32E-08	2.66E-09	1.249	1.16-1.35
6	rs3104398	HLA-DQA2	32,793,663	A	0.11	8.16E-08	2.81E-07	0.718	0.64-0.81
6	rs2261033	BAT2	31,711,570	C	0.49	9.83E-08	2.49E-06	1.219	1.14-1.31
6	rs204999	PRRT1	32,217,957	G	0.26	1.13E-07	2.30E-09	0.795	0.73-0.86
6	rs9277554	HLA-DPB1	33,163,516	T	0.297	1.33E-07	4.64E-08	1.236	1.15-1.33
6	rs2516398	MICB	31,589,505	C	0.312	1.53E-07	1.64E-07	0.807	0.75-0.87
6	rs6901221	HLA-DPB2	33,206,254	C	0.166	1.91E-07	1.19E-07	1.29	1.18-1.41
6	rs7774954	HLA-DQB2	32,832,167	A	0.066	2.13E-07	3.06E-07	0.662	0.57-0.77
6	rs2248462	MICB	31,554,775	A	0.207	2.22E-07	3.27E-07	0.784	0.72-0.86
6	rs2596480	MICA	31,533,964	A	0.077	2.40E-07	6.55E-07	0.683	0.59-0.79
6	rs707939	MSH5	31,834,667	T	0.37	2.43E-07	6.55E-07	1.217	1.13-1.31
6	rs2516509	MICB	31,557,973	G	0.207	3.08E-07	4.19E-07	0.786	0.72-0.86
6	rs2075800	HSPA1L	31,885,925	A	0.355	3.70E-07	1.58E-07	1.216	1.13-1.31
6	rs3095352	DDR1	30,913,900	G	0.424	4.19E-07	3.12E-06	0.826	0.77-0.89
6	rs12665700	C6orf205	31,104,111	T	0.137	8.18E-07	7.25E-07	1.293	1.17-1.43
6	rs3129941	C6orf10	32,445,664	A	0.227	9.50E-07	1.43E-06	0.801	0.74-0.87
6	rs6941112	STK19	32,054,593	A	0.343	1.24E-06	2.29E-07	1.206	1.12-1.30

**Supplementary Table 3. Association analysis of overall top SNPs on each population separately. Right most column show Breslow-Day P value for the heterogeneity on the odds ratio.**

CHR	Gene	SNP	Position	Minor Allele	Population	MAF (case/control)	P Value*	OR (95 CI)	Breslow-day P value
6p21	HLA-DQB1	rs6457617	32771829	C	United States	0.403/0.487	1.17E-14	0.71 (0.65-0.77)	5.11E-02
				C	Spain	0.402/0.474	5.22E-03	0.74 (0.61-0.92)	
				C	Germany	0.400/0.504	4.67E-05	0.66 (0.54-0.80)	
				C	Holland	0.491/0.488	9.14E-01	1.01 (0.80-1.28)	
7q32	TNPO3/ IRF5	rs10488631	128381419	C	United States	0.149/0.105	5.44E-10	1.49 (1.13-1.69)	4.34E-01
				C	Spain	0.115/0.090	1.03E-01	1.32 (0.94-1.85)	
				C	Germany	0.170/0.101	2.62E-05	1.84 (1.38-2.44)	
				C	Holland	0.125/0.095	9.69E-02	1.36 (0.94-1.97)	
		A	United States	0.169/0.132	1.27E-06	1.34 (1.19-1.51)	5.80E-01		
		A	Spain	0.143/0.132	5.24E-01	1.10 (0.82-1.48)			
		A	Germany	0.161/0.125	3.94E-02	1.34 (1.01-1.78)			
		A	Holland	0.145/0.117	1.52E-01	1.28 (0.91-1.80)			
2q32	STAT4	rs3821236	191611003	A	United States	0.234/0.204	7.74E-04	1.19 (1.08-1.32)	3.97E-02
				A	Spain	0.278/0.189	4.88E-05	1.65 (1.29-2.10)	
				A	Germany	0.257/0.194	2.54E-03	1.44 (1.13-1.82)	
				A	Holland	0.278/0.203	2.39E-03	1.52 (1.16-1.99)	
1q22-23	CD247	rs2056626	165687049	G	United States	0.379/0.428	5.82E-06	0.82 (0.75-0.89)	4.26E-01
				G	Spain	0.324/0.384	1.60E-02	0.77 (0.62-0.95)	
				G	Germany	0.344/0.402	1.95E-02	0.78 (0.63-0.96)	
				G	Holland	0.420/0.425	8.90E-01	0.98 (0.77-1.25)	
6p25	EXOC2/ IRF4	rs4959270	402748	A	United States	0.466/0.502	1.33E-03	0.87 (0.80-0.95)	5.88E-02
				A	Spain	0.408/0.513	4.63E-05	0.65 (0.53-0.80)	
				A	Germany	0.392/0.466	3.54E-03	0.74 (0.60-0.91)	
				A	Holland	0.426/0.470	1.39E-01	0.84 (0.66-1.06)	
18q22	CDH7	rs10515998	61521202	G	United States	0.056/0.039	1.32E-04	1.47 (1.20-1.79)	5.30E-03
				G	Spain	0.060/0.059	9.65E-01	1.01 (0.66-1.55)	
				G	Germany	0.093/0.052	1.20E-03	1.86 (1.27-2.72)	
				G	Holland	0.062/0.018	1.28E-05	3.54 (1.94-6.48)	

\* P values uncorrected for  $\lambda$

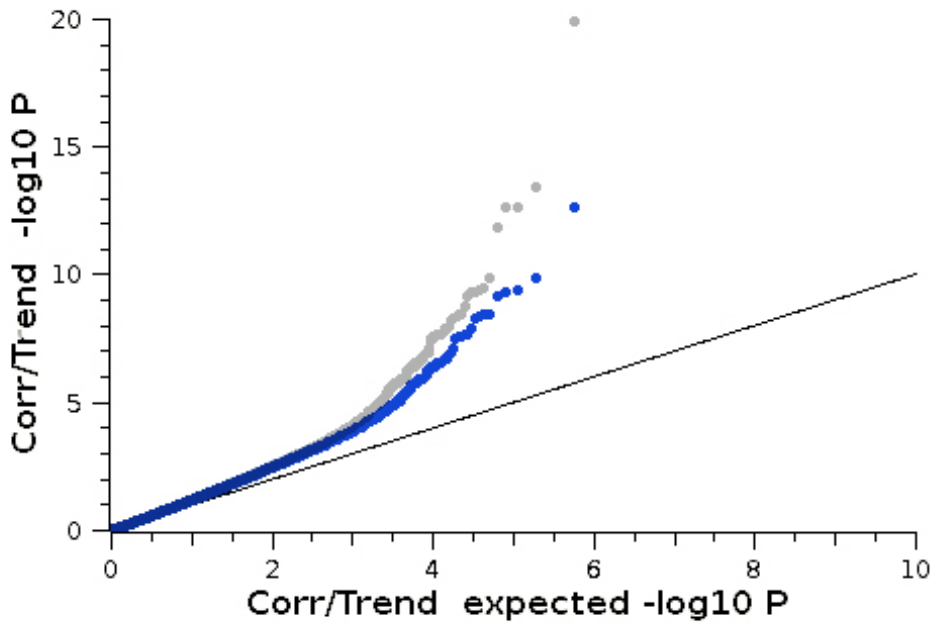
**Supplementary table 4. Replication phase study design.**

<b>Population</b>	<b>N Case/Controls</b>
Belgium	189/274
Spain	455/739
The Netherlands	228/279
Germany	207/285
Italy	348/727
Norway	113/283
Sweden	279/455
UK	500/384
US	434/1143
<b>Total</b>	<b>2753/4569</b>

**Supplementary table 5. Main clinical features of SSc patients included in the study.**

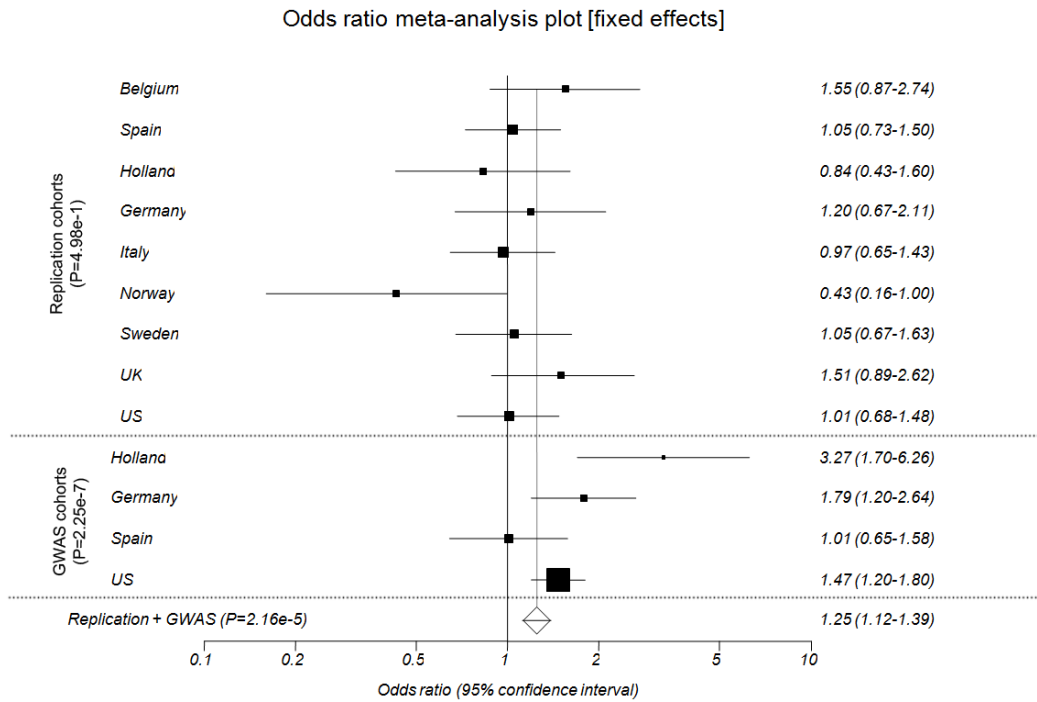
	Sex (cases/controls)		Subtype		Anti-centromere		Anti-topoisomerase	
	Female	Male	Difuse	Limited	Positive	Negative	Positive	Negative
Overall	0.80/0.79	0.11/0.21	0.31	0.58	0.32	0.58	0.20	0.71
US	0.88/0.88	0.12/0.12	0.34	0.61	0.29	0.62	0.16	0.77
Spain	0.88/0.75	0.10/0.25	0.25	0.59	0.42	0.41	0.20	0.61
German	0.88/0.62	0.11/0.38	0.35	0.44	0.42	0.53	0.31	0.62
Dutch	0.72/0.51	0.28/0.49	0.13	0.51	0.22	0.69	0.26	0.65
<b>Replication SSc cohorts</b>								
Belgium	0.73/0.45	0.21/0.54	0.31	0.63	0.23	0.47	0.18	0.52
Spain	0.87/0.60	0.11/0.37	0.32	0.59	0.29	0.61	0.14	0.75
The Netherlands	0.71/0.45	0.20/0.54	0.17	0.51	0.22	0.62	0.28	0.64
Germany	0.81/0.44	0.13/0.27	0.37	0.45	0.27	0.35	0.21	0.41
Italy	0.76/0.65	0.07/0.35	0.20	0.60	0.31	0.53	0.32	0.52
Norway	0.85/0.33	0.15/0.66	0.35	0.65	0.53	0.46	0.14	0.84
Sweden	0.78/0.78	0.22/0.22	0.28	0.72	0.27	0.73	0.17	0.83
UK	0.83/0.44	0.16/0.56	0.27	0.72	0.28	0.54	0.11	0.70
US	0.88/0.44	0.12/0.56	0.37	0.50	0.32	0.66	0.12	0.85

**Supplementary figure 1. Quantile-quantile (QQ) plot of the observed P values for association.** The grey dots represent the totality of SNPs analyzed ( $\lambda = 1.069$ ) and the blue dots represent all the SNPs analyzed excluding the MHC region ( $\lambda = 1.066$ ).

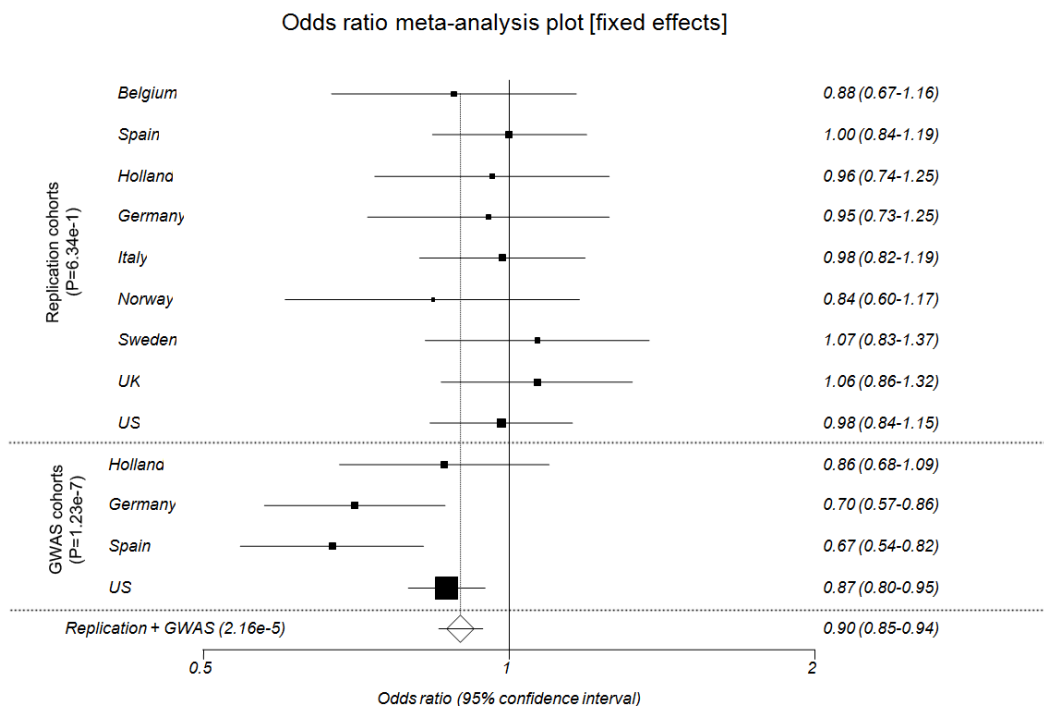


**Supplementary figure 2. Forest plot of the *CDH7* and *EXOC2/IRF4* loci at different study stages.**

**A) *CDH7***

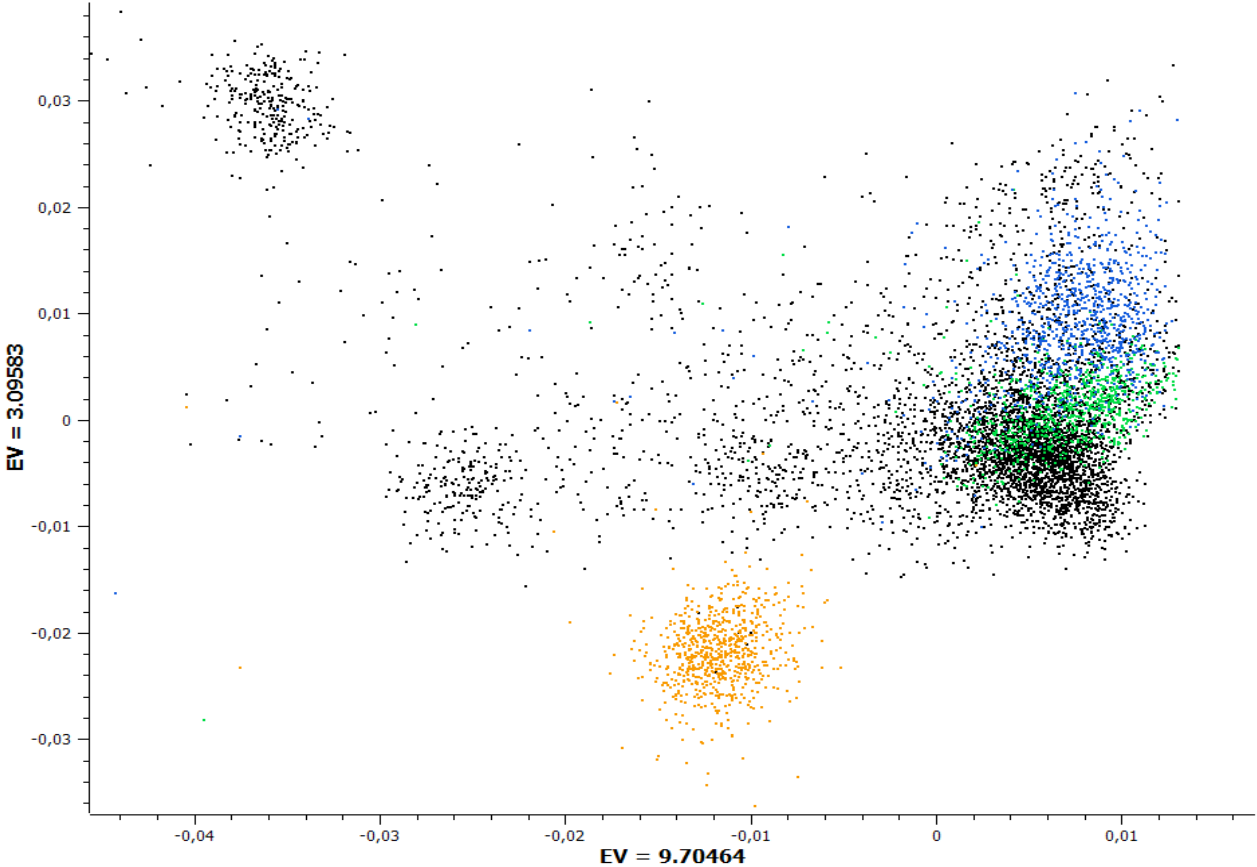


**B) *EXOC2/IRF4***

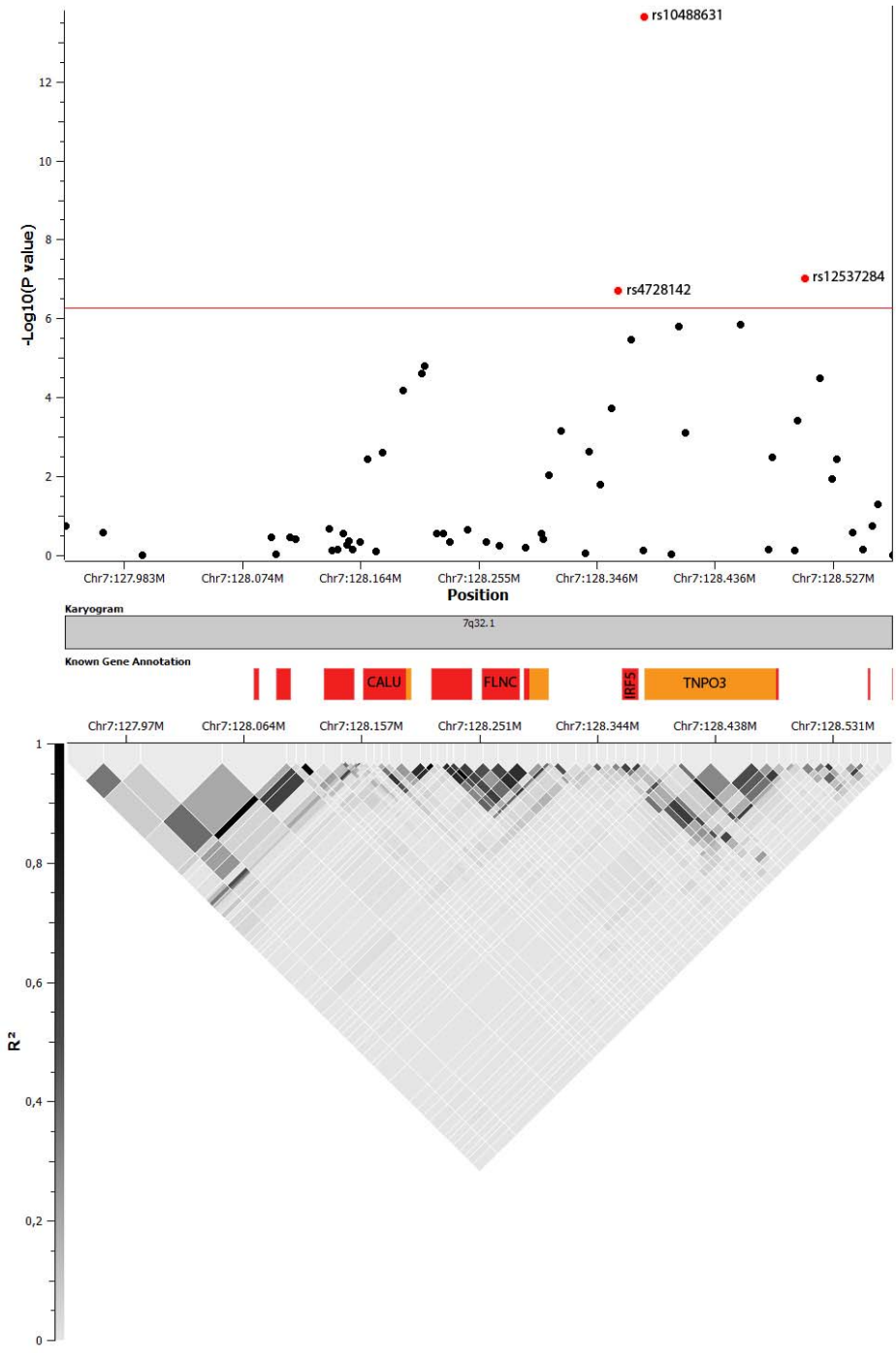




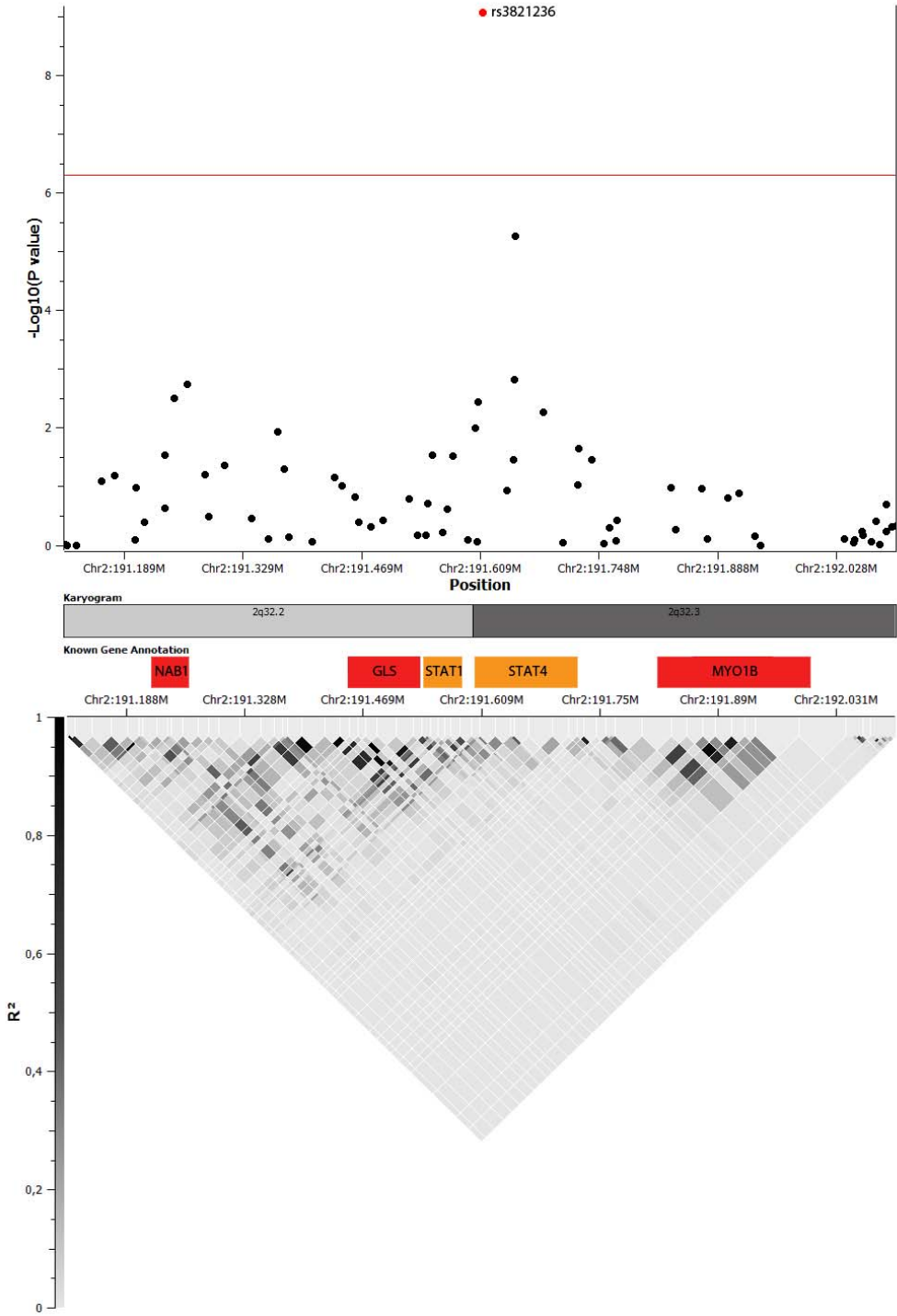
**Supplementary figure 3.** Principal component plot for the first 2 eigenvectors. US population is represented on black, German on blue, Dutch on green and Spanish on orange.



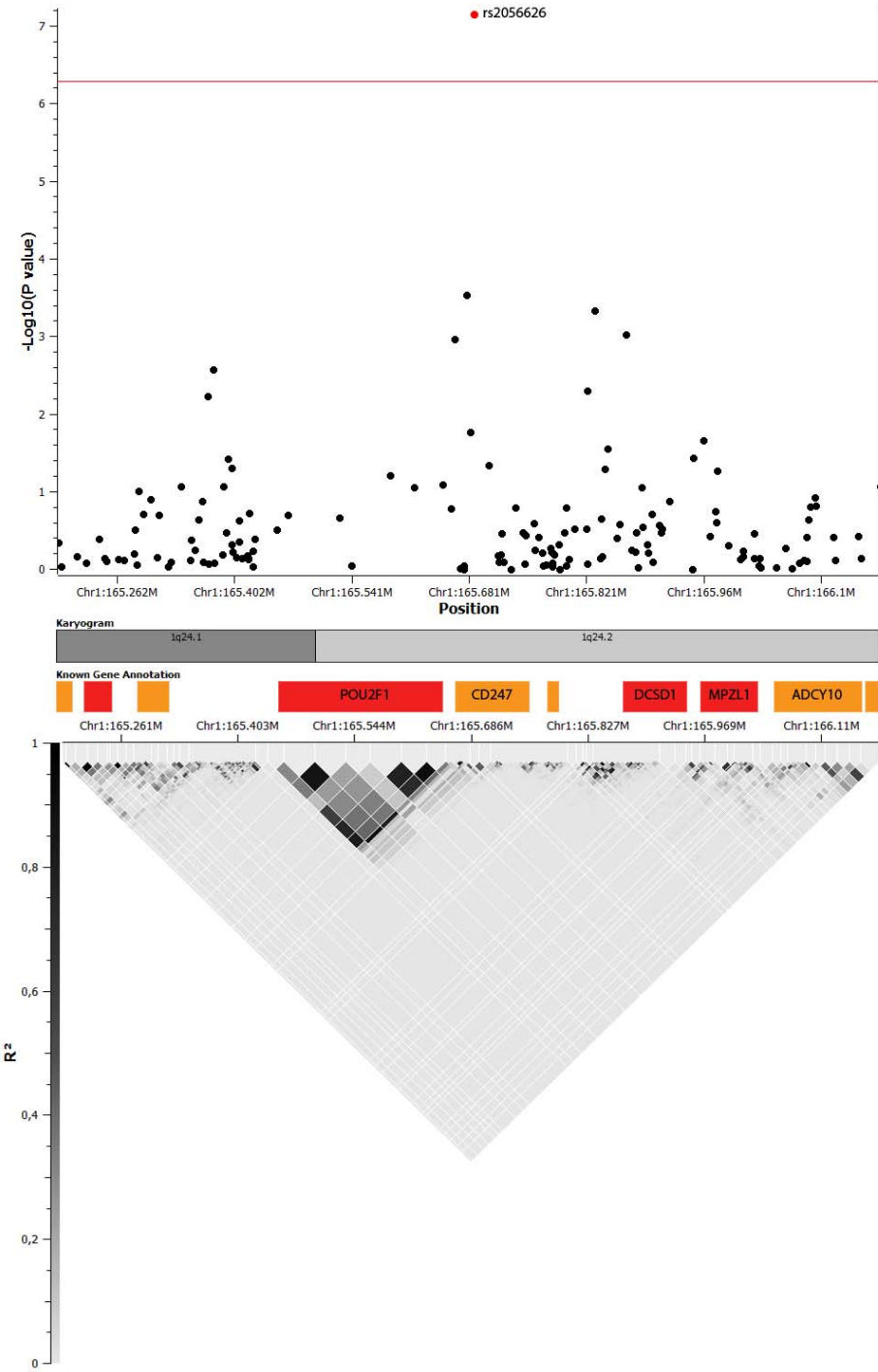
**Supplementary figure 4a.** Association and linkage disequilibrium (LD) plot on TNPO3/IRF5 association region. P values are uncorrected for  $\lambda$ . LD represented is  $R^2$ . Genome wide significance level is marked with a red line.



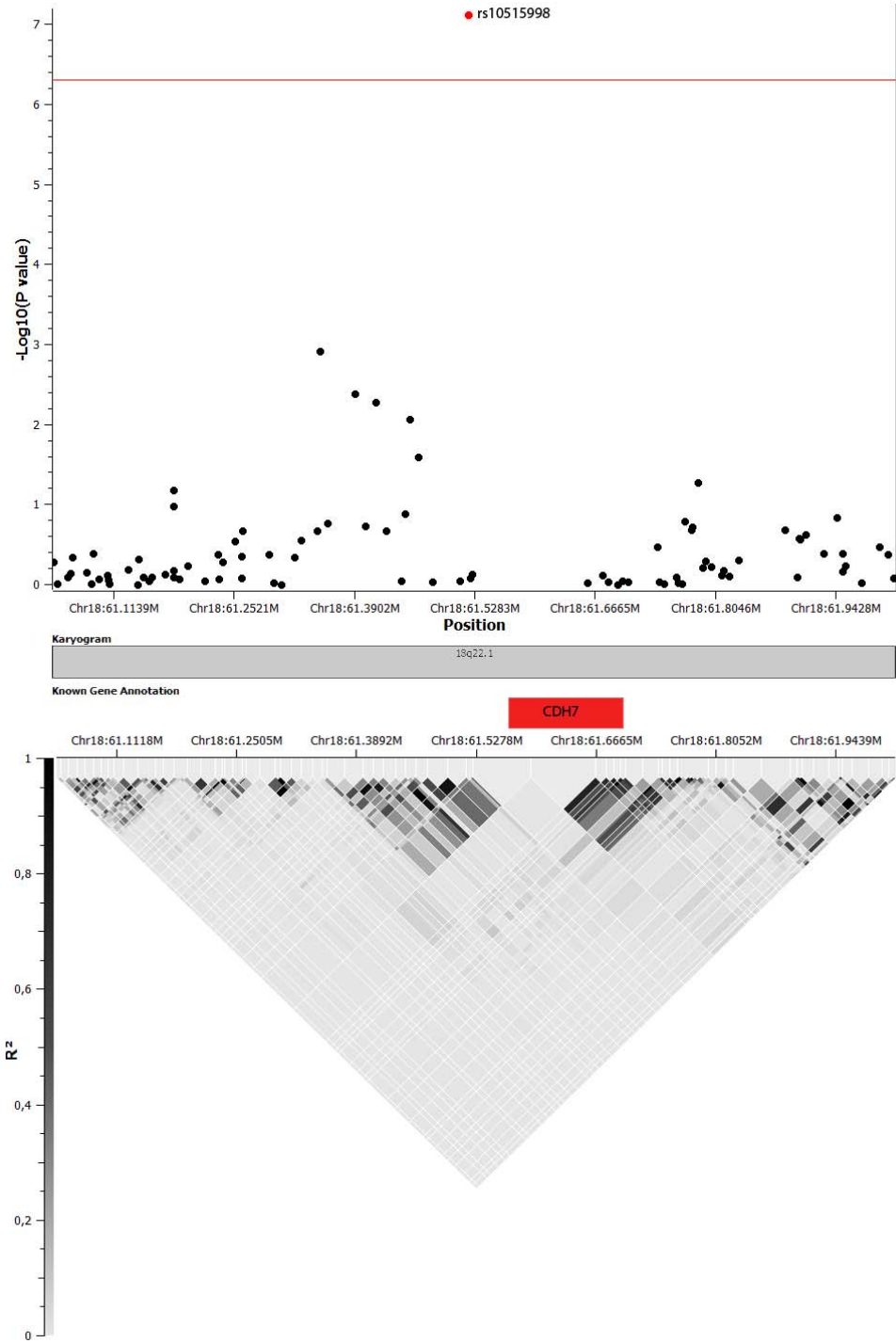
**Supplementary figure 4b.** Association and linkage disequilibrium (LD) plot on STAT4 association region. P values are uncorrected for  $\lambda$ . LD represented is  $R^2$ . Genome wide significance level is marked with a red line.



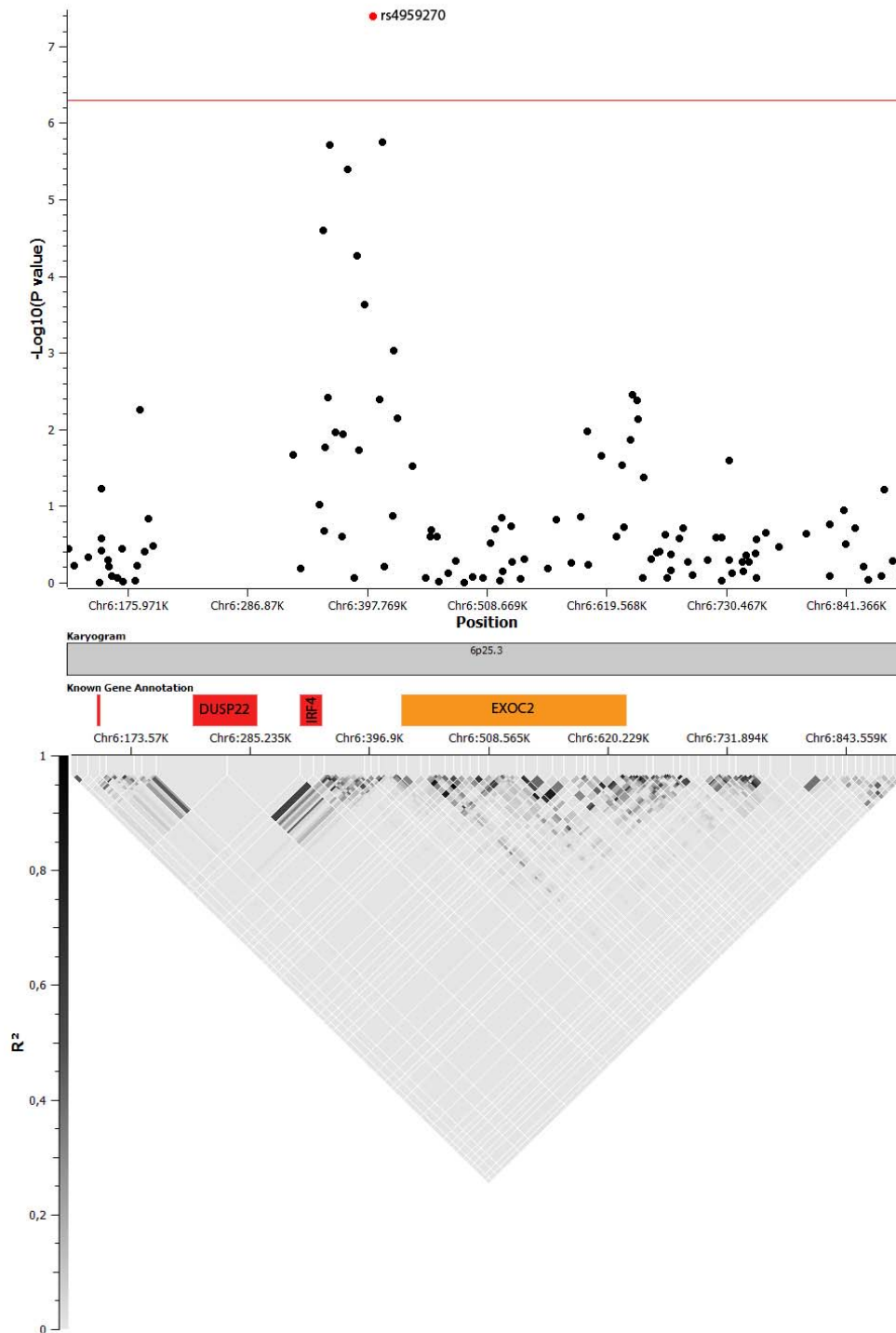
**Supplementary figure 4c.** Association and linkage disequilibrium (LD) plot on CD247 association region. P values are uncorrected for  $\lambda$ . LD represented is  $R^2$ . Genome wide significance level is marked with a red line.



**Supplementary figure 4d.** Association and linkage disequilibrium (LD) plot on CDH7 association region. P values are uncorrected for  $\lambda$ . LD represented is  $R^2$ . Genome wide significance level is marked with a red line.



**Supplementary figure 4e.** Association and linkage disequilibrium (LD) plot on EXOC2/IRF4 association region. P values are uncorrected for  $\lambda$ . LD represented is  $R^2$ . Genome wide significance level is marked with a red line.



## ***Supplementary note***

### **Spanish SSc Group**

Vicente Fonollosa, Hospital Valle de Hebrón, Barcelona; Mayte Camps and Antonio Fernández-Nebro, Hospital Carlos Haya, Málaga; Rosa García-Portales, Hospital Virgen de la Victoria, Málaga; M<sup>a</sup> Angeles Aguirre-Zamorano, Hospital Reina Sofía, Córdoba; Francisco J García-Hernández and Julio Sánchez-Roman, Hospital Virgen del Rocío, Anna Pros, Hospital del Mar, Barcelona; Luis Rodriguez, Hospital Clínico San Carlos, Madrid; Esther Vicente, Hospital La Princesa, Madrid; Jose Roman-Iborra, Hospital Dr. Peset, Valencia; Bernardino Díaz, Hospital Central de Asturias, Oviedo; Paloma García de la Peña, Hospital Universitario Ramón y Cajal, Madrid.