Supplementary file 1 – Literature and meta-analyses

A literature search was performed to investigate associations between HLA-SE, smoking and anti-citrullinated protein antibodies (ACPA) in the asymptomatic general population and first-degree relatives (FDR). See supplementary figure 1 for a flowchart of the article selection. PubMed was searched until September 2020 with terms as: *"SE"[All Fields] AND ("smoke"[MeSH Terms] OR "smoke"[All Fields] OR "smoke s"[All Fields] OR "smoked"[All Fields] OR "smokes"[All Fields] OR "smoking"[MeSH Terms] OR "smoking"[All Fields] OR "smokings"[All Fields] OR "smoking s"[All Fields]) AND ("autoantibodies"[MeSH Terms] OR "autoantibodies"[All Fields] OR "autoantibody"[All Fields]) AND ("healthies"[All Fields] OR "healthy"[All Fields])*, including different combinations of SE, shared epitope, smoking, autoantibodies, ACPA, healthy, asymptomatic, preclinical, general population, population-based and FDR (see below for all combinations used). Additional articles were identified by hand searching reference lists. After removal of reviews, duplicates and articles that did not apply to our research question (based on title and abstract screening), ten articles remained. Subsequently all studies that were either cross-sectional or longitudinal studies and contained information on associations between HLA-SE, smoking and ACPA-development were considered eligible. No longitudinal studies starting the healthy population were identified; all identified studies were cross-sectional in nature. However, since ACPA-development is most likely the first event in development of ACPA-positive disease, we believe the observed findings in cross-sectional studies reflect effects of HLA-SE and smoking on ACPA-development. The ten selected articles were further studied on the description of the population. Studies evaluating populations in which part of subjects had swollen and/or tender joints, and studies in which associations with autoantibodies were not investigated for ACPA (separate from other autoantibodies) were excluded. Finally, six cross-sectional studies were eligible for meta-analyses. Four studies evaluated associations between HLA-SE and ACPA [1-4], and five studies evaluated associations between smoking and ACPA. [1,2,4-6] Meta-analyses were first performed without a stringent evaluation of study bias and thereafter with assessing this bias as recommended according to the PRISMA guidelines.[7] First, the odds ratios (OR) of these four and five studies respectively were

combined in inverse-variance weighted meta-analyses, I^2 was determined to evaluate heterogeneity across the studies. Some studies were truly population based (no selection), these were the studies of van Zanten et al. (2017) and Terao et al. (2014). Other studies included asymptomatic persons whom were selected because of having a relative with RA [3,4], being a twin [2], or were blood bank donors.[6] In the latter no odds ratio was given on the association of smoking and ACPA in healthy individuals, however, since all required information was available we calculated the odds ratio from the data presented in the publication. In Terao et al. (2014) results were stratified for gender; for the present meta-analyses we included results presented for women, however, results from the meta-analysis was similar when men were included in the analyses (data not shown).

Next, to eliminate possible inclusion bias, only the two studies that evaluated unselected individuals from the general population (van Zanten et al. (2017) and Terao et al. (2014)) were included in the meta-analysis. Then similar results were obtained, see supplementary figure 2.

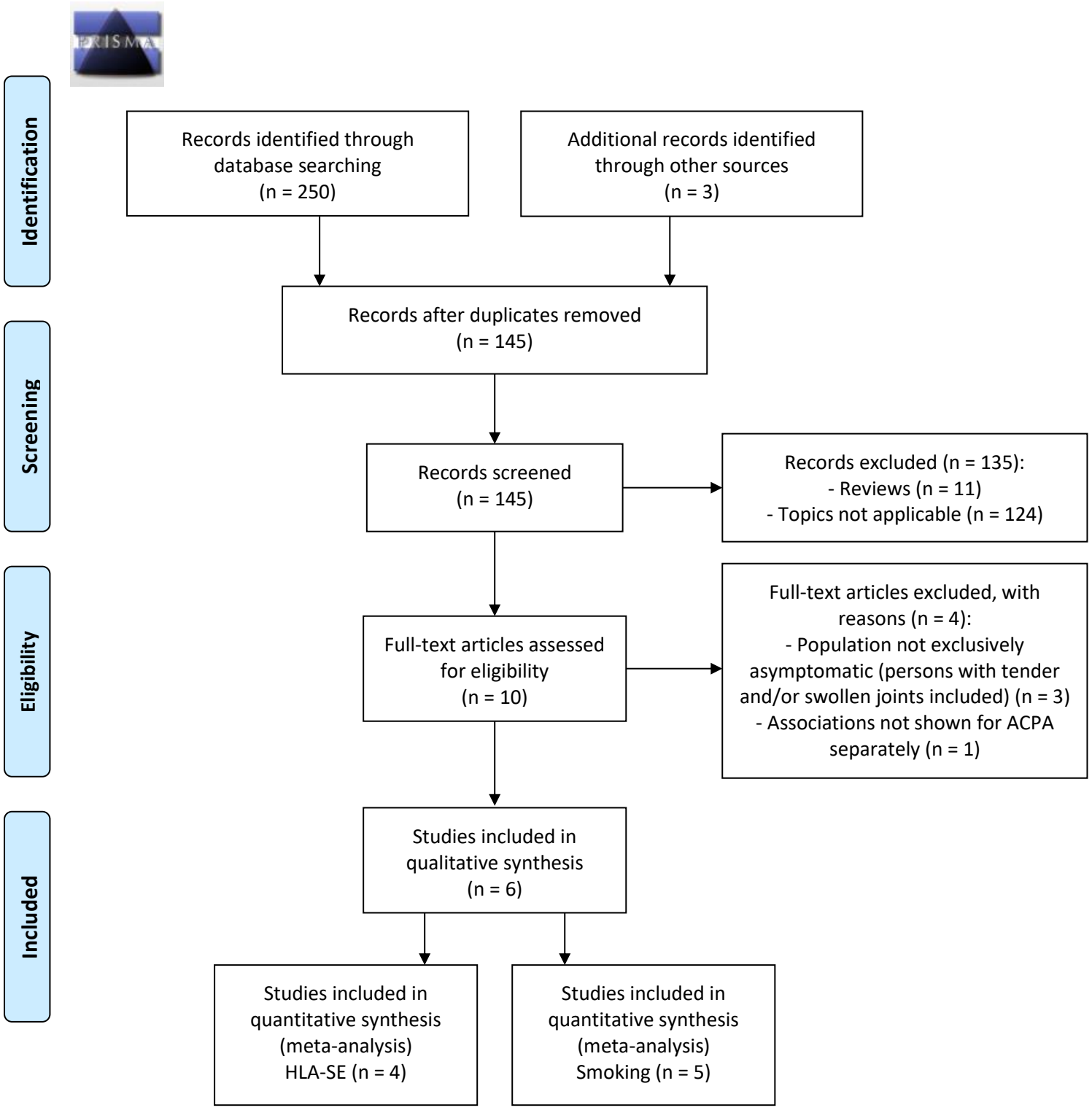
Finally, since recent data in RA-patients indicated that smoking does not associate with ACPA as such, but rather with RF or autoantibodies in general [8-12], and we wished to evaluate whether this is the same in the general population, we retrieved RF data, if present, from the identified studies, and performed additional meta-analyses. Two studies (Terao et al. (2014) and Ärlestig et al. (2011)) provided data on associations of HLA-SE and smoking with RF. Meta-analyses indicated that was no association was present between HLA-SE and RF (OR 0.97 (95% CI 0.70-1.35)), nor between smoking and RF (OR 0.84 (0.55-1.28); forest plots not shown).

List of search terms that were used:

- SE smoking autoantibodies healthy (15 results)
- SE smoking autoantibodies asymptomatic (0 results)
- SE smoking autoantibodies preclinical (0 results)
- SE smoking ACPA healthy (8 results)
- SE ACPA healthy (32 results)
- shared epitope ACPA healthy (33 results)

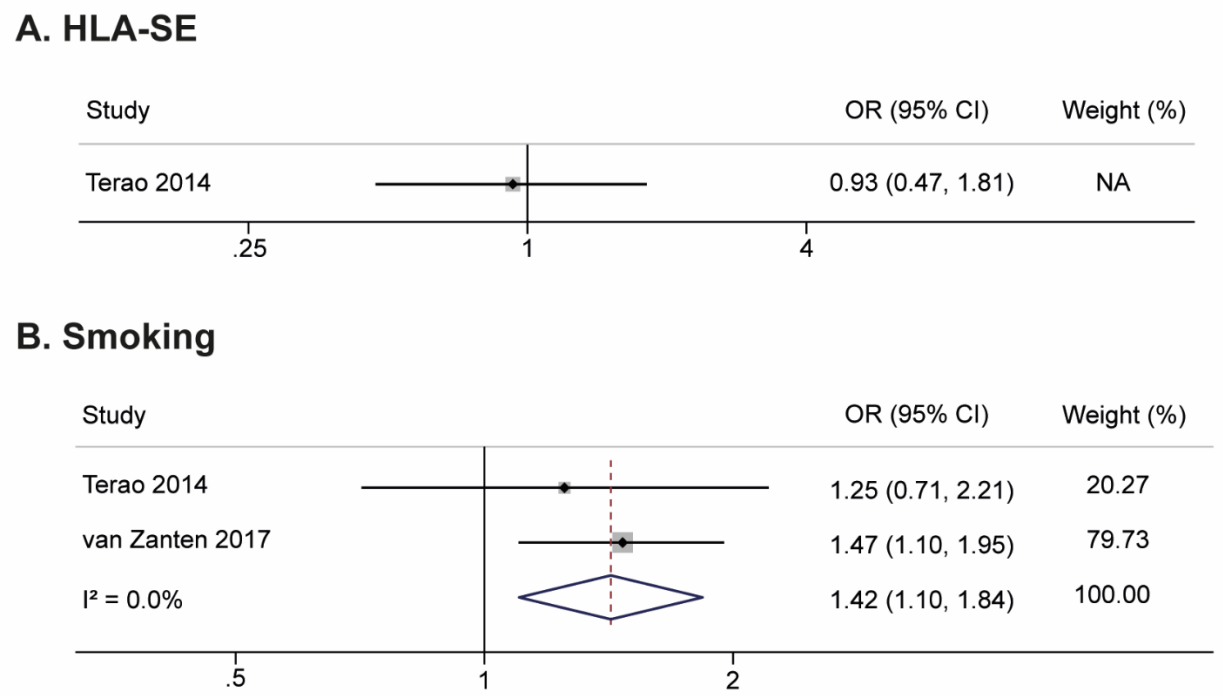
- smoking ACPA healthy (32 results)
- smoking ACPA general population (16 results)
- shared epitope ACPA general population (10 results)
- smoking ACPA population-based (25 results)
- shared epitope ACPA population-based (15 results)
- smoking ACPA FDR (2 results)
- SE ACPA FDR (2 results)
- shared epitope ACPA FDR (2 results)
- SE smoking ACPA (58 results)

Supplementary figure 1 Flowchart of literature review and article selection for meta-analyses on association of HLA-SE and smoking in the asymptomatic population.



Meta-analyses were performed in the 4 (HLA-SE) en 5 (smoking) included studies, respectively, without assessment of inclusion bias. Some of these studies did not select persons from the general asymptomatic population. Therefore sensitivity analyses with the population-based studies were performed, see supplementary figure 2.

Supplementary figure 2 Meta-analyses on HLA-SE (A) and smoking (B) with presence of ACPA in unselected asymptomatic healthy population-based individuals



HLA-SE: shared epitope, ACPA: anti-citrullinated protein antibody, OR: odds ratio, CI: confidence interval

Supplementary file 2 – Detailed description of methods

Leiden CSA cohort

Patients presenting with CSA to the Leiden rheumatology outpatient clinic between April 2012-September 2019 were studied. The cohort is described in detail previously.[13] Patients had recent-onset (<1 year) arthralgia of small joints and were, according to the clinical expertise and pattern recognition of the rheumatologist, at risk for progression to RA. Autoantibody status was largely unknown at inclusion as (in line with Dutch guidelines) general practitioners in the area of Leiden are discouraged to perform autoantibody tests. Inclusion in the CSA-cohort was therefore predominantly based on history taking and physical examination. Patients were excluded if clinical arthritis was already present, or if a different explanation for the joint pain was more likely. Patients were followed for at least 2 years for development of clinically apparent inflammatory arthritis (IA), defined as ≥1 swollen joints at physical examination by a rheumatologist, with scheduled follow-up visits at 4, 12 and 24 months. In case of an increase in symptoms or suspicion of clinical arthritis additional visits were planned. During follow-up treatment with disease-modifying anti-rheumatic drugs (DMARDs, including systemic or intra-articular corticosteroids) was not allowed, therefore patients participating in a randomized controlled trial investigating the effect of methotrexate in the phase of CSA (RCT; treat earlier) were excluded from the current study.

Baseline visit consisted of physical examination, blood sampling, questionnaires (including questions on current and past smoking) and an MRI. Presence of IgM RF (in house ELISA, cut-off >3.5 IU/mL) and IgG ACPA (anti-CCP2, Phadia, Nieuwegein, the Netherlands, cut-off>7 U/mL) was determined during routine laboratory measurements in all patients. Because of limited laboratory capacity not all patients were selected for additional anti-CarP and AAPA measurements. Patient selection for these tests was first based on availability of samples and presence of RF and/or ACPA at baseline. In total 189 patients were included for additional autoantibody measurements. First, patients positive for RF and/or ACPA at baseline with available serum samples were included (n=89, 45 progressing and 44 non-progressing patients). Then, autoantibody-negative patients with available samples that

progressed to IA were additionally included (n=37). Finally, from the large group of autoantibody-negative patients that did not progress a random selection was made (n=63). Baseline characteristics of the randomly selected non-progressing autoantibody-negative patients were similar to that of the patients that were not selected (data not shown); suggesting that the selection is representative for this total group. The fact that anti-CarP and AAPA were only measured in a selection of autoantibody-negative patients limits interpretation of predictive accuracy of these autoantibodies in the entire CSA-population. However, since no selection was made in RF-positive/ACPA-positive patients, analyses were repeated and validated with the Amsterdam cohort in this subgroup.

The study was approved by the local medical ethical committee and all participants gave written informed consent.

HLA-SE

HLA-SE status was derived from whole genome sequencing data. All patients were whole genome sequenced using the Illumina Global Screening Array (GSA). Standard quality control steps were performed using Plink v1.90 [14]; individuals with a missingness of more than 2% were removed, as were variants with a missingness of over 2%, a minor allele frequency (MAF) below 0.01 or a Hardy-Weinberg equilibrium p-value below 0.0001. In order to prepare the genotype data for imputation of the HLA region we used the McCarthy program [15], which checks for ambiguous SNPs with a MAF above 0.4 (only A/T & G/C SNPs), SNPs with a MAF deviating more than 0.2 from the HRC reference panel [16] and SNPs not in the aforementioned reference panel. We subsequently use the SNP2HLA tool with the T1DGC reference to impute the HLA region.[17] From these imputed data we extracted the HLA-SE, defined as the HLA-DRB1 variants *0101, *0102, *0401, *0404, *0405, *0408 and *1001.[18] Patients were deemed HLA-SE positive if 1 or 2 of the SE variants were present.

Anti-carbamylated and anti-acetylated antibodies

In serum, we determined the presence and levels of IgG anti-carbamylated and anti-acetylated protein antibodies (anti-CarP and AAPA, respectively). In-house ELISA was used for all measurements as described previously [19]. Briefly, plates were coated with carbamylated FCS and CCP1 acetylated lysine for measurements of anti-CarP and AAPA, respectively. To determine background signal, plates were additionally coated with non-modified antigens (non-modified FCS and CCP1 norleucine, respectively). Serum samples were diluted 1:50 and incubated. After washing, plates were subsequently incubated with HRP-labeled rabbit-anti-human IgG (Dako). HRP-activity was visualized with ABTS and measurements were expressed in arbitrary units per milliliter (aU/mL). On every plate a dilution standard was included to determine the linear part of the curve; standards from all plates were used in the analyses. The fourth standard, which is expected to be in the middle (and therefore linear part) of the curve, is further diluted and additionally included as a reference sample. Serum of healthy subjects (n=199) was used to determine the cut-off of all autoantibody measurements, which was calculated as the mean plus two times the standard deviation of healthy subjects. When the difference in optical density (OD) between the non-modified antigens and the modified proteins was <0.1, the measurement was considered non-specific; non-specific measurements with values above the cut-off were considered negative. Inter-assay variation of in-house ELISAs was determined previously by reevaluation of ~10% of samples; measurements were highly correlated (Pearson's r ranges 0.96-0.97) and changes in positivity of the test were infrequent, see supplementary figure 4. Intra-assay variability was determined for anti-CarP by measurement of 3 samples 10 times, the mean coefficient of variation (CV, mean % (SD)) was 20.4 (6.8). Of note, although not absolute at the monoclonal- or polyclonal level, cross-reactivity of ACPA towards other post translationally modified proteins have been conclusively shown in different studies [20,21], and hence should be regarded as anti-modified protein antibody-reactivities.

Amsterdam arthralgia cohort

Analyses evaluating progression to IA were stratified for ACPA-status and results from the ACPA-positive subgroup were studied in meta-analyses with the results from two other ACPA-positive

cohorts. The first cohort concerned arthralgia patients from Amsterdam, as described in van de Stadt et al. [22]. As this study evaluated ACPA- and/or RF-positive patients, raw data from ACPA-positive individuals was obtained for this study. Data on smoking was obtained by history taking. Presence of HLA-SE was inferred from HLA-DQA1, HLA-DQB1 haplotypes using strong linkage disequilibrium with HLA-DRB1 alleles in Caucasians.[23] Patients were positive for HLA-SE if 1 or 2 copies of the HLA-DRB1*0101, *0102, *0401, *0404, *0405, *0408, *0410 or *1001 alleles were present. Presence of IgM RF and ACPA were determined with in-house ELISA (cut-off >30 IU/mL) and aCCP ELISA (Axis Shield; cut-off>5 aU/mL), respectively. Presence of anti-CarP was determined with in-house ELISA similar to methods used in the CSA cohort.[19] Baseline serum samples were obtained to determine presence of IgG AAPA simultaneously with Leiden CSA-samples.

Leeds ACPA-positive cohort

The second cohort consists of ACPA-positive patients with non-specific musculoskeletal symptoms from the Leeds cohort, as described by Rakieh et al. [24]. Since only ACPA-positive patients were studied, results obtained from the literature were sufficient for inclusion in the meta-analyses; raw data was not evaluated. Also here smoking status was obtained through history taking. Presence of HLA-SE was determined by PCR amplification with sequence-specific primers.[25] Patients were positive for HLA-SE if 1 or 2 copies of alleles in the HLA-DRB1*01, *04 and *10 locus were present. Presence of IgM RF (initial cut-off>40 IU/mL, later >20 IU/mL) and ACPA (anti-CCP2, immunocap assay, Phadia; initial cut-off>7 IU/mL, later >2.99 IU/mL) was determined. No serum samples were obtained from this cohort and the presence of anti-CarP and AAPA was not determined.

REFERENCES

[1] Terao C, Ohmura K, Ikari K, Kawaguchi T, Takahashi M, Setoh K et al. Effects of smoking and shared epitope on the production of anti-citrullinated peptide antibody in a Japanese adult population. *Arthritis Care Res (Hoboken)* 2014;66:1818-27.

[2] Hensvold AH, Magnusson PK, Joshua V, Hansson M, Israelsson L, Ferreira R et al. Environmental and genetic factors in the development of anticitrullinated protein antibodies (ACPAs) and ACPA-positive rheumatoid arthritis: an epidemiological investigation in twins. *Ann Rheum Dis* 2015;74:375-80.

[3] El-Gabalawy HS, Robinson DB, Hart D, Elias B, Markland J, Peschken CA et al. Immunogenetic risks of anti-cyclical citrullinated peptide antibodies in a North American Native population with rheumatoid arthritis and their first-degree relatives. *J Rheumatol* 2009;36:1130-5.

[4] Arlestig L, Mullazehi M, Kokkonen H, Rocklov J, Ronnelid J, Dahlqvist SR. Antibodies against cyclic citrullinated peptides of IgG, IgA and IgM isotype and rheumatoid factor of IgM and IgA isotype are increased in unaffected members of multicase rheumatoid arthritis families from northern Sweden. *Ann Rheum Dis* 2012;71:825-9.

[5] van Zanten A, Arends S, Roozendaal C, Limburg PC, Maas F, Trouw LA et al. Presence of anticitrullinated protein antibodies in a large population-based cohort from the Netherlands. *Ann Rheum Dis* 2017;76:1184-90.

[6] Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum* 2006;54:38-46.

[7] Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 2010;8:336-41.

[8] van Wesemael TJ, Ajeganova S, Humphreys J, Terao C, Muhammad A, Symmons DP et al. Smoking is associated with the concurrent presence of multiple autoantibodies in rheumatoid arthritis rather than with anti-citrullinated protein antibodies per se: a multicenter cohort study. *Arthritis Res Ther* 2016;18:285.

[9] Ishikawa Y, Ikari K, Hashimoto M, Ohmura K, Tanaka M, Ito H et al. Shared epitope defines distinct associations of cigarette smoking with levels of anticitrullinated protein antibody and rheumatoid factor. *Ann Rheum Dis* 2019;78:1480-7.

[10] Murphy D, Mattey D, Hutchinson D. Anti-citrullinated protein antibody positive rheumatoid arthritis is primarily determined by rheumatoid factor titre and the shared epitope rather than smoking per se. *PLoS One* 2017;12:e0180655.

[11] Hedstrom AK, Ronnelid J, Klareskog L, Alfredsson L. Complex Relationships of Smoking, HLA-DRB1 Genes, and Serologic Profiles in Patients With Early Rheumatoid Arthritis: Update From a Swedish Population-Based Case-Control Study. *Arthritis Rheumatol* 2019;71:1504-11.

[12] Regueiro C, Rodriguez-Rodriguez L, Lopez-Mejias R, Nuno L, Triguero-Martinez A, Perez-Pampin E et al. A predominant involvement of the triple seropositive patients and others with rheumatoid factor in the association of smoking with rheumatoid arthritis. *Sci Rep* 2020;10:3355.

[13] van Steenbergen HW, van Nies JA, Huizinga TW, Bloem JL, Reijnierse M, van der Helm-van Mil AH. Characterising arthralgia in the preclinical phase of rheumatoid arthritis using MRI. *Ann Rheum Dis* 2015;74:1225-32.

[14] Plink v1.90. Available from: https://www.cog-genomics.org/plink/.

[15] McCarthy Program. Available from: https://www.well.ox.ac.uk/~wrayner/tools/.

[16] HRC reference panel. Available from: http://www.haplotype-reference-consortium.org/site.

[17] Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* 2013;8:e64683.

[18] van der Woude D, Lie BA, Lundstrom E, Balsa A, Feitsma AL, Houwing-Duistermaat JJ et al. Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. *Arthritis Rheum* 2010;62:1236-45.

[19] Shi J, Knevel R, Suwannalai P, van der Linden MP, Janssen GM, van Veelen PA et al. Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. *Proc Natl Acad Sci U S A* 2011;108:17372-7.

[20] Kissel T, Reijm S, Slot LM, Cavallari M, Wortel CM, Vergroesen RD et al. Antibodies and B cells recognising citrullinated proteins display a broad cross-reactivity towards other post-translational modifications. *Ann Rheum Dis* 2020;79:472-80.

[21] Sahlstrom P, Hansson M, Steen J, Amara K, Titcombe PJ, Forsstrom B et al. Different Hierarchies of Anti-Modified Protein Autoantibody Reactivities in Rheumatoid Arthritis. *Arthritis Rheumatol* 2020.

[22] van de Stadt LA, Witte BI, Bos WH, van Schaardenburg D. A prediction rule for the development of arthritis in seropositive arthralgia patients. *Ann Rheum Dis* 2013;72:1920-6.

[23] Knipper AJ, Hakenberg P, Enczmann J, Kuhrober A, Kiesel U, Kogler G et al. HLA-DRB1,3,4,5 and -DQB1 allele frequencies and HLA-DR/DQ linkage disequilibrium of 231 German caucasoid patients and their corresponding 821 potential unrelated stem cell transplants. *Hum Immunol* 2000;61:605-14.

[24] Rakieh C, Nam JL, Hunt L, Hensor EM, Das S, Bissell LA et al. Predicting the development of clinical arthritis in anti-CCP positive individuals with non-specific musculoskeletal symptoms: a prospective observational cohort study. *Ann Rheum Dis* 2015;74:1659-66.

[25] Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue Antigens* 1992;39:225-35.

Supplementary table 1 Baseline characteristics

	CSA cohort Leiden			Arthralgia cohort Amsterdam
	All patients (n=577)	*ACPA-positive (n=77)*	*ACPA-negative (n=500)*	*ACPA-positive (n=244)*
Female, n (%)	451 (78.2)	61 (79.2)	390 (78.0)	186 (76.2)
Age in years, mean (SD)	43.7 (12.6)	48.1 (12.0)	43.0 (12.5)	48.4 (11.4)
Symptom duration in weeks, median (IQR)	20 (9-46)	25 (13-53)	19 (9-43)	52 (28-104)
53-TJC, median (IQR)	5 (2-10)	3 (2-7)	5 (2-10)	0 (0-3)
Increased CRP, n (%)	126 (22.0)	23 (30.3)	103 (20.7)	35 (14.5)
RF positivity, n (%)	113 (19.6)	59 (76.6)	54 (10.8)	108 (44.3)
ACPA positivity, n (%)	77 (13.3)	NA	NA	NA

Data on HLA-SE and smoking was available in the CSA cohort in 522 and 500 patients, respectively. In Amsterdam in 133 and 243 patients, respectively.

CSA: clinically suspect arthralgia, ACPA: anti-citrullinated protein antibody, SD: standard deviation, IQR: interquartile range, TJC: tender joint count, RF: rheumatoid factor, CRP: c-reactive protein

Supplementary table 2 Associations of HLA-SE and smoking with ACPA-positivity stratified for RF, in patients newly presenting with CSA

		ACPA positive, n (%)	ACPA negative, n (%)	OR (95% CI)	p-value
RF-positive patients					
HLA-SE	Absent	22 (42)	32 (64)	Reference	--
	Present	31 (59)	18 (36)	2.51 (1.13-5.55)	0.024
HLA-SE	0	22 (42)	32 (64)	Reference	--
	1	23 (43)	16 (32)	2.09 (0.91-4.83)	0.084
	2	8 (15)	2 (4)	5.82 (1.13-30.05)	0.036
Smoking	Never	13 (26)	23 (49)	Reference	--
	Ever	37 (74)	24 (51)	2.73 (1.16-6.40)	0.021
Smoking	Never	13 (26)	23 (49)	Reference	--
	Ex-smoker	22 (44)	21 (45)	1.85 (0.75-4.58)	0.18
	Current smoker	15 (30)	3 (6)	8.85 (2.15-36.38)	0.003
RF-negative patients					
HLA-SE	Absent	5 (31)	227 (56)	Reference	--
	Present	11 (69)	176 (44)	2.84 (0.97-8.32)	0.057
HLA-SE	0	5 (31)	227 (56)	Reference	--
	1	8 (50)	145 (36)	2.51 (0.80-7.81)	0.11
	2	3 (19)	31 (8)	4.39 (1.0-19.30)	0.050
Smoking	Never	2 (14)	162 (42)	Reference	--
	Ever	12 (86)	227 (58)	4.28 (0.95-19.39)	0.059
Smoking	Never	2 (14)	162 (42)	Reference	--
	Ex-smoker	6 (43)	140 (36)	3.47 (0.69-17.48)	0.13
	Current smoker	6 (43)	87 (22)	5.59 (1.10-28.27)	0.038

HLA-SE: shared epitope, ACPA: anti-citrullinated protein antibody, RF: rheumatoid factor, CSA: clinically suspect arthralgia, OR: odds ratio, CI: confidence interval

Supplementary table 3 Associations of HLA-SE and smoking with ACPA-level in ACPA-positive patients evaluated with Mann-Whitney U tests and logistic regression

HLA-SE positive	**HLA-SE negative**	**Mann-Whitney U**	**Logistic regression**
Median ACPA level (IQR)	*Median ACPA level (IQR)*	*p-value*	*OR (95% CI), p-value*
236 (72-340)	144 (32-340)	0.12	1.002 (0.999-1.006), 0.18

Ever smoking	**Never smoking**	**Mann-Whitney U**	**Logistic regression**
Median ACPA level (IQR)	*Median ACPA level (IQR)*	*p-value*	*OR (95% CI), p-value*
229 (64-340)	222 (52-340)	0.89	1.000 (0.996-1.004), 0.99

Odds ratios indicate effect measures for every unit increase in ACPA-level.

HLA-SE: shared epitope, ACPA: anti-citrullinated protein antibody, IQR: interquartile range, OR: odds ratio, CI: confidence interval

Supplementary table 4 Associations of HLA-SE with development of inflammatory arthritis in patients newly presenting with CSA; hazard ratios corresponding to Figure 2 of the main article

		IA, n (%)	No IA, n (%)	HR (95% CI)	p-value
All patients					
HLA-SE	Absent	39 (42)	247 (58)	Reference	--
	Present	55 (59)	181 (42)	1.86 (1.23-2.82)	0.003
HLA-SE	0	39 (42)	247 (58)	Reference	--
	1	41 (44)	151 (35)	1.65 (1.06-2.56)	0.027
	2	14 (15)	30 (7)	3.03 (1.64-5.61)	<0.001
ACPA positive subgroup					
HLA-SE	Absent	15 (37)	12 (43)	Reference	--
	Present	26 (63)	16 (57)	1.29 (0.67-2.47)	0.44
HLA-SE	0	15 (37)	12 (43)	Reference	--
	1	17 (42)	14 (50)	1.05 (0.52-2.13)	0.90
	2	9 (22)	2 (7)	2.32 (1.00-5.41)	0.051
ACPA negative subgroup					
HLA-SE	Absent	24 (45)	235 (59)	Reference	--
	Present	29 (55)	165 (41)	1.71 (0.99-2.96)	0.055
HLA-SE	0	24 (45)	235 (59)	Reference	--
	1	24 (45)	137 (34)	1.66 (0.94-2.94)	0.083
	2	5 (9)	28 (7)	2.00 (0.76-5.28)	0.16

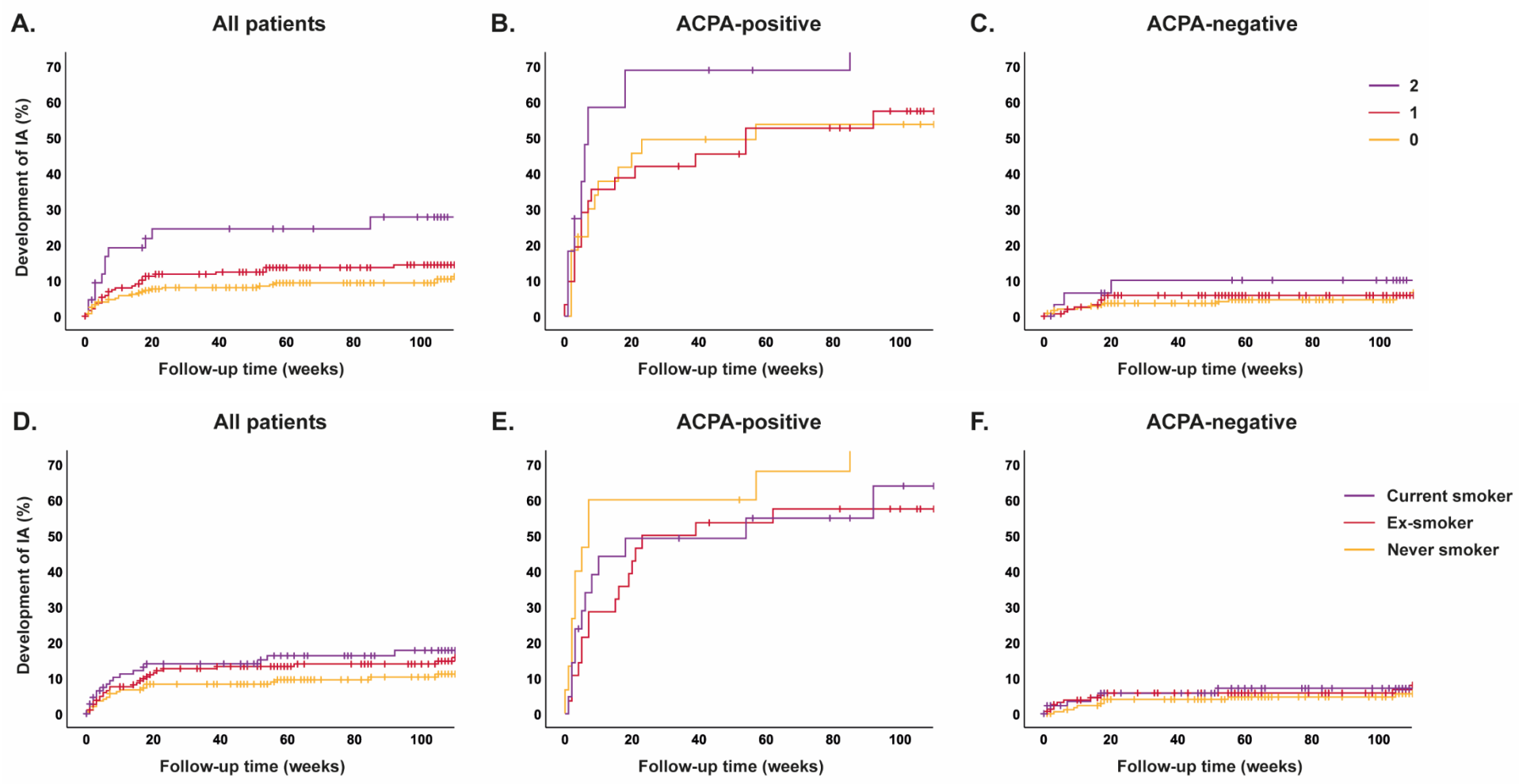
HLA-SE: shared epitope, CSA: clinically suspect arthralgia, IA: inflammatory arthritis, HR: hazard ratio, CI: confidence interval, ACPA: anti-citrullinated protein antibody

Supplementary table 5 Associations of smoking with development of inflammatory arthritis in patients newly presenting with CSA; hazard ratios corresponding to Figure 3 of the main article

		IA, n (%)	No IA, n (%)	HR (95% CI)	p-value
All patients					
Smoking	Never	29 (32)	171 (42)	Reference	--
	Ever	63 (69)	237 (58)	1.40 (0.90-2.18)	0.14
Smoking	Never	29 (32)	171 (42)	Reference	--
	Ex-smoker	38 (41)	151 (37)	1.25 (0.76-2.06)	0.37
	Current smoker	25 (27)	86 (21)	1.66 (0.97-2.83)	0.065
ACPA positive subgroup					
Smoking	Never	11 (28)	4 (17)	Reference	--
	Ever	29 (73)	20 (83)	0.59 (0.29-1.18)	0.13
Smoking	Never	11 (28)	4 (17)	Reference	--
	Ex-smoker	17 (43)	11 (46)	0.55 (0.26-1.19)	0.13
	Current smoker	12 (30)	9 (38)	0.64 (0.28-1.45)	0.28
ACPA negative subgroup					
Smoking	Never	18 (35)	167 (44)	Reference	--
	Ever	34 (65)	217 (57)	1.30 (0.73-2.33)	0.37
Smoking	Never	18 (35)	167 (44)	Reference	--
	Ex-smoker	21 (40)	140 (37)	1.17 (0.61-2.24)	0.64
	Current smoker	13 (25)	77 (20)	1.56 (0.76-3.18)	0.23

CSA: clinically suspect arthralgia, IA: inflammatory arthritis, HR: hazard ratio, CI: confidence interval, ACPA: anti-citrullinated protein antibody

Supplementary figure 3 Associations of HLA-SE (A-C) and smoking (D-F) with progression from CSA to RA



RA is defined as fulfilment of the 1987 and/or 2010 criteria at the time of clinically apparent inflammatory arthritis development.

Corresponding hazard ratios, with 0 HLA-SE alleles as reference category were: (A) HR 1.42 (95% CI 0.85-2.38) and HR 2.97 (1.48-5.96) for 1 and 2 HLA-SE alleles respectively, (B) HR 1.05 (0.52-2.13) and HR 2.08 (0.87-5.00), and (C) HR 1.16 (0.53-2.53) and HR 1.87 (0.54-6.51).

Corresponding hazard ratios, with never smoker as reference category were: (D) HR 1.44 (0.81-2.56) and HR 1.81 (0.97-3.39) for ex-smoker and current smoker respectively, (E) HR 0.55 (0.26-1.19) and HR 0.64 (0.28-1.45), and (F) HR 1.42 (0.59-3.43) and HR 1.66 (0.62-4.46).

HLA-SE: shared epitope, CSA: clinically suspect arthralgia, RA: rheumatoid arthritis, ACPA: anti-citrullinated protein antibody, HR: hazard ratio, CI: confidence interval

Supplementary table 6 Associations of HLA-SE and smoking with presence of RF, anti-CarP and AAPA in patients newly presenting with CSA

RF	Univariable OR (95% CI)	p-value	Multivariable [a] OR (95% CI)	p-value
HLA-SE present	1.13 (0.73-1.73)	0.59	0.71 (0.41-1.22)	0.21
Ever-smoker	1.16 (0.74-1.84)	0.52	0.70 (0.40-1.23)	0.22

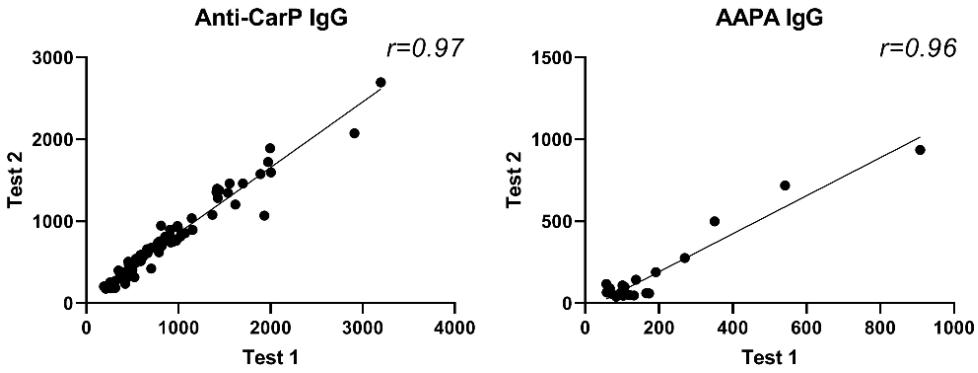
Anti-CarP	Univariable OR (95% CI)	p-value	Multivariable [a] OR (95% CI)	p-value
HLA-SE present	1.63 (0.59-4.50)	0.34	1.31 (0.42-4.07)	0.65
Ever-smoker	1.05 (0.37-2.95)	0.93	0.58 (0.17-2.01)	0.39

AAPA	Univariable OR (95% CI)	p-value	Multivariable [a] OR (95% CI)	p-value
HLA-SE present	1.41 (0.65-3.05)	0.39	1.15 (0.47-2.82)	0.76
Ever-smoker	0.58 (0.26-1.31)	0.19	0.28 (0.098-0.80)	0.018

[a] Corrected for presence of ACPA

HLA-SE: shared epitope, RF: rheumatoid factor, anti-CarP: anti-carbamylated protein antibody, AAPA: anti-acetylated protein antibody, CSA: clinically suspect arthralgia, OR: odds ratio, CI: confidence interval, ACPA: anti-citrullinated protein antibody

Supplementary figure 4 Inter-assay variation of in-house ELISAs



Inter-assay variation resulted in changes in positivity of the test infrequently: anti-CarP IgG 3.9%, AAPA IgG 4.2%.

anti-CarP: anti-carbamylated protein antibody, AAPA: anti-acetylated protein antibody