OPEN

A predominant involvement of the triple seropositive patients and others with rheumatoid factor in the association of smoking with rheumatoid arthritis

Cristina Regueiro[1], Luis Rodriguez-Rodriguez [2], Raquel Lopez-Mejias [3], Laura Nuño[4], Ana Triguero-Martinez[5], Eva Perez-Pampin[1], Alfonso Corrales[3], Alejandro Villalba[4], Yolanda Lopez-Golan[1], Lydia Abasolo[2], Sara Remuzgo-Martínez[3], Ana M. Ortiz[5], Eva Herranz[2], Ana Martínez-Feito[6], Carmen Conde [1], Antonio Mera-Varela[1,7], Alejandro Balsa[4], Isidoro Gonzalez-Alvaro[5], Miguel Ángel González-Gay[3], Benjamín Fernandez-Gutierrez[2] & Antonio Gonzalez [1]*

The major environmental risk factor for rheumatoid arthritis (RA) is smoking, which according to a widely accepted model induces protein citrullination in the lungs, triggering the production of anti-citrullinated protein antibodies (ACPA) and RA development. Nevertheless, some research findings do not fit this model. Therefore, we obtained six independent cohorts with 2253 RA patients for a detailed analysis of the association between smoking and RA autoantibodies. Our results showed a predominant association of smoking with the concurrent presence of the three antibodies: rheumatoid factor (RF), ACPA and anti-carbamylated protein antibodies (ACarPA) (3 Ab *vs.* 0 Ab: OR = 1.99, p = 2.5 × 10^{-8}). Meta-analysis with previous data (4491 patients) confirmed the predominant association with the concurrent presence of the three antibodies (3 Ab *vs.* 0 Ab: OR = 2.00, p = 4.4 × 10^{-16}) and revealed that smoking was exclusively associated with the presence of RF in patients with one or two antibodies (RF^{+}_{1+2} *vs.* RF^{-}_{0+1+2}: OR = 1.32, p = 0.0002). In contrast, no specific association with ACPA or ACarPA was found. Therefore, these results showed the need to understand how smoking favors the concordance of RA specific antibodies and RF triggering, perhaps involving smoking-induced epitope spreading and other hypothesized mechanisms.

Rheumatoid arthritis (RA) is a systemic autoimmune disease that can be divided in two pathogenic subgroups[1,2]. The largest subgroup comprises the patients presenting RA specific autoantibodies. These antibodies include the rheumatoid factor (RF), which is directed against the Fc of the IgG, and antibodies against some post-translational protein modifications. The best characterized are anti-citrullinated protein antibodies (ACPA), which in the clinic are assayed as anti-cyclic citrullinated peptides or anti-CCP, and the anti-carbamylated protein antibodies (ACarPA), which are not yet analyzed beyond research studies. The antibody positive patients are known as seropositive and they represent more than two thirds of the total, while the remaining are the seronegative patients.

[1]Experimental and Observational Rheumatology and Rheumatology Unit, Instituto de Investigacion Sanitaria, Hospital Clínico Universitario de Santiago (IDIS), Santiago de Compostela, Spain. [2]Rheumatology Department, Hospital Clínico San Carlos, Instituto Investigación Sanitaria San Carlos (IdISSC), Madrid, Spain. [3]Valdecilla Biomedical Research Institute, Hospital Universitario Marqués de Valdecilla (IDIVAL), Santander, Spain. [4]Rheumatology Department, Instituto de Investigación Hospital Universitario La Paz (IDIPAZ), Madrid, Spain. [5]Rheumatology Department, Instituto de Investigación Sanitaria la Princesa, Hospital Universitario de la Princesa (IIS-IP), Madrid, Spain. [6]Immuno-Rheumatology Department, Instituto de Investigación Hospital Universitario La Paz (IDIPAZ), Madrid, Spain. [7]Faculty of Medicine and Dentistry, University of Santiago de Compostela, Santiago de Compostela, Spain. *email: agmartinezp@ser.es

The seropositive patients have a clear component of genetic susceptibility and a defined disease story and autoimmune pathogenesis, which are less clear in the seronegative patients[1–5].

Also, smoking, which is the major environmental risk factor for RA, is specific of seropositive RA[6,7]. In the ACPA positive patients, the risk is potentiated by interaction with the HLA-DRB1 shared epitope (SE)[8–10]. This interaction together with protein citrullination in the bronchoalveolar lavage (BAL) cells of smokers has been at the basis of an influential pathogenic model[11]. According to this model, smoking induces protein citrullination in the lung, which predisposes to the production of ACPA and, subsequently, of RA[11]. This model has been supported by other findings, including the demonstration of increased PAD2 expression in bronchoalveolar lavage (BAL) of smokers[12,13]. However, not all the findings have been consistent. In particular, no correlation between smoking and protein citrullination[12,14,15] or between smoking and PAD2[15] have been observed in the lung tissue specimens. Moreover, the production of ACPA in the lung and the presence of lung abnormalities in early RA were directly associated with ACPA$^+$ RA without detectable association with smoking in the available studies[16–19]. Also, van Wesemael et al. reported that smoking was associated with the concurrent presence of RF, ACPA and ACarPA, rather than with ACPA in three patient cohorts[20]. The novelty and repercussion of these latter results ask for validation through independent replication.

Here, we have replicated and extended the findings of van Wesemael et al. in an independent set of six patient cohorts (n = 2253). Our results confirmed the predominant association of smoking with the concurrent presence of the three RA autoantibodies. Also, thanks to the large power afforded by the combined analysis (n = 4491), we discovered that smoking was exclusively associated with RF positivity in the patients without three concordant positive autoantibodies (RF$^+_{1+2}$). In contrast, we did not find significant associations of smoking with the specific presence of ACPA or ACarPA. These results indicate that smoking promotes pathways leading to the concurrent presence of the three RA autoantibodies and, in its defect, to the production of RF.

Material and Methods

Patients and samples. Patients from five Spanish (IDIPAZ, PEARL, IDIS, IdISSC, and IDIVAL) and one Italian (Rome) RA cohorts were considered as replication sets (n = 2253 with complete data). The Spanish data were directly available to us, whereas the information corresponding to the Italian patients was extracted from a publication[21]. Two of the replication collections were early arthritis (EA) prospective clinics (IDIPAZ and PEARL), whereas the remaining four replication cohorts (IDIS, IdISSC, IDIVAL and Rome) were from established RA patients. Entry criteria for IDIPAZ[22] and PEARL[23] were: 2 or more swollen joints for less than a year and absence of previous treatment with Disease Modifying Anti-Rheumatic Drugs (DMARD). In addition, the patients with RA according to the 1987 ACR criteria[24] at 2 years of follow-up and with serum samples and smoking information in the baseline visit (IDIPAZ, n = 243; and PEARL, n = 264) were selected. The established Spanish RA patients (according to 1987 ACR criteria, with serum samples and information on smoking) were from IDIS[25], n = 470, IdISSC[26], n = 508, and IDIVAL[27], n = 459. In turn, the Italian patients (n = 309) were classified according to the 2010 ACR criteria[21]. Smoking information was obtained as never smoker, past smoker or current smoker in response to the written questionnaire given to the patients at recruitment, either at the first visit (all patients in IDIPAZ and PEARL) or at any time (in the remaining cohorts). No information on smoking intensity was available from most patients. All the patients included in this study granted their written informed consent. The study was designed and conducted according to the Declaration of Helsinki, the Belmont Report and the Spanish Law 14/2007 of Biomedical Research. This included the approval of the Hospital Clínico San Carlos Ethics Committee, the Cantabria Ethics Committee, the EAC by the La Paz University Hospital Ethics Committee, the Ethics Committee for Clinical Research of Hospital Universitario La Princesa (Ref. PI-518), and the approval of the study by the Autonomous Research Ethics Committee of Galicia (Ref. 2014/387 and 2017/514).

For meta-analysis, data from the three patient collections (NOAR, EAC Leiden and BARFOT) included in the van Wesemael study (n = 2238 with complete information) were retrieved[20].

Anti-CarP antibodies and other RA autoantibodies. Anti-CarP IgG antibodies were assessed by ELISA as previously described[25,28]. IgM-RF was determined by rate nephelometry, whereas ACPA were determined by ELISA. The ACPA were tested as anti-CCP2 with the EDIA enzyme-linked immunosorbent assay kit in all the IDIS and IDIVAL patients and with the Immunoscan RA in all the IDIPAZ and IdISSC patients (both assays from Euro Diagnostica, Malmö, Sweden). The patients of PEARL were tested with Immunoscan RA until October 2010. Afterward, they were assayed with the QUANTA Lite CCP3 IgG and IgA assay (Inova Diagnostics, San Diego, CA).

Statistical analysis. Results from the different patient cohorts, and from previously reported cohorts[20,21], were combined by meta-analysis with the R package *meta*[29]. Subgroup meta-analysis comparing EAC and prevalent RA cohorts was done with Review Manager Version 5.3[30]. Smoking habit was considered as ever or never smoker status. In most analyses, the RF$^-$ACPA$^-$ACarPA$^-$ triple negative patients (0 Ab) were used as the reference. Alternative meta-analyses compared other patient subgroups, which are indicated in the text with a combination of the antibody abbreviation (RF, ACPA or ACarPA), the plus and minus superscripts for presence/absence, and the number of antibodies considered (from 0 to 3) as subscripts. In this way, the patients bearing only RF are coded as RF$^+_1$, those bearing RF and a second antibody as RF$^+_2$. See Supplementary Table S1 for the remaining codes. Heterogeneity between cohorts was assessed with the inconsistency parameter I^2. By default, the fixed effects model was applied for meta-analysis, weighting the contribution of each cohort with the inverse variance method. The random effects model according to DerSimonian and Laird was preferred when heterogeneity was notable ($I^2 > 50$) and reported as OR$_{re}$ and p$_{re}$. P values lower than 0.05 were considered statistically significant.

	New cohorts						van Wesemael *et al.*		
Cohort	IDIPAZ[a]	PEARL[a]	IDIS	IdISSC	IDIVAL	Rome[b]	NOAR[a]	EAC Leiden[a]	BARFOT[a]
Complete data, n	243	264	470	508	459	309	674	769	795
Ever smoker, n (%)	117 (48.1)	115 (43.6)	96 (20.4)	232 (45.7)	253 (55.1)	167 (54)	432 (63.7)	415 (54)	471 (59.2)
RF positive, n (%)	168 (69.1)	175 (66.3)	281 (59.8)	329 (64.8)	235 (51.2)	202 (65.4)	277 (40.9)	442 (57.5)	448 (56.4)
ACPA positive, n (%)	170 (70)	163 (61.7)	300 (63.8)	251 (49.4)	225 (49)	190 (61.5)	247 (36.4)	404 (52.5)	456 (57.4)
ACarPA positive, n (%)	109 (44.9)	100 (37.9)	141 (30)	158 (31.1)	149 (32.5)	117 (37.9)	182 (26.8)	349 (45.4)	279 (35.1)
Number of autoAb.									
0	54 (22.2)	59 (22.3)	111 (23.6)	141 (27.8)	173 (37.7)	62 (20.1)	292 (43.1)	242 (31.5)	263 (33.1)
1	21 (8.6)	45 (17)	95 (20.2)	118 (23.2)	78 (17)	67 (21.7)	167 (24.6)	129 (16.8)	110 (13.8)
2	78 (32.1)	87 (33)	165 (35.1)	127 (25)	93 (20.3)	98 (31.7)	108 (15.9)	128 (16.6)	193 (24.3)
3	90 (37)	73 (27.7)	99 (21.1)	122 (24)	115 (25.1)	82 (26.5)	107 (15.8)	270 (35.1)	229 (28.8)

Table 1. Main characteristics of the patients with RA. Analysis of the six new cohorts (n = 2253) that were studied here for the first time and of the three cohorts (n = 2238) from a previous study[20]. [a]Early arthritis patients. [b]Data extracted from Pecani *et al.*[21].

Additionally, we used exploratory analysis on the pooled data across the cohorts for interpretation of the findings. It included graphic representation and classification trees. For the former, we employed a double-decker plot[31]. The classification trees, in turn, were done with the General Classification and Regression Trees module of Statistica (v7.0, StatSoft, Tulsa, OK) that produces an exhaustive and recursive search of the best classification. We considered nine dichotomous variables to classify the patients according to smoking. The variables represented the presence or absence of the antibodies, their combinations and number: RF, ACPA, ACarPA, RF&ACPA, RF&ACarPA, ACPA&ACarPA, one antibody, two antibodies, and three antibodies. The priors were considered proportional to the smoking class sizes, the misclassification costs were taken to be equal for every class, the splits were selected based in the Gini index of node impurity, and no stops were imposed. This procedure searches the minimum number of univariate splits to produce the tree with less misclassified patients without being limited by the high dependence between the classification variables.

Results

Replication of the association of smoking with concurrent autoantibodies.
The six newly obtained cohorts of RA patients included 2253 patients with complete data (Table 1). Two of the cohorts were prospective EAC (n = 507), the remaining included patients with established RA (n = 1746). About half (43.6–55.1%) of the patients were ever smokers, except in one of the cohorts where the frequency of smokers was notably lower (20.4%). This circumstance is characteristic of the population attending the recruiting hospital, as previously noted[32]. The other critical characteristic for this study, the presence of autoantibodies, varied between the cohorts within the commonly observed range: mean percentages of $RF^+ = 61.7\%$, $ACPA^+ = 57.7\%$, and $ACarPA^+ = 34.4\%$. The patients were divided into four subgroups by the number of antibodies they presented. These subgroups were of comparable size except for the patients with one antibody that were the less abundant: 26.6, 18.8, 28.8 and 25.8% for the groups with 0, 1, 2 and 3 antibodies, respectively.

We assessed the relationship of smoking with the patients grouped according to the number of antibodies they presented (Table 2). Only the patients positive for the three autoantibodies were significantly associated with smoking (3 Ab *vs.* 0 Ab: OR = 1.99, p = 2.5×10^{-8}). In stark contrast with this very significant result, the patients with one or two autoantibodies were not significantly different from the patients without antibodies (Table 2). In addition, the association of smoking with the concurrent presence of the three antibodies was significant not only relative to the patients without antibodies (Table 2), but also relative to the patients with one antibody (3 Ab *vs.* 1 Ab:OR = 1.68, 95% CI = 1.32–2.13, p = 2.2×10^{-5}, $I^2 = 0\%$) and with two antibodies (3 Ab *vs.* 2 Ab: OR = 1.61, 95% CI = 1.27–2.04, p = 8.1×10^{-5}, $I^2 = 0\%$).

Combined meta-analysis of the available data.
We combined the 6 replication cohorts with the 3 cohorts form van Wesemael *et al.*[20] to a total of 4491 patients with RA. The summary data showed that the association with the concurrent presence of the three antibodies was highly significant (Fig. 1C, 3 Ab *vs.* 0 Ab: OR = 2.00, p = 4.4×10^{-16}, $I^2 = 17\%$), and significantly stronger than the observed with the patients carrying two antibodies when directly compared (3 Ab *vs.* 2 Ab: OR = 1.54, 95% CI 1.29–1.84, p = 1.4×10^{-6}, $I^2 = 12\%$). Even so, the patients with two concordant positive antibodies were associated with smoking (Fig. 2B, 2 Ab *vs.* 0 Ab: OR = 1.26, p = 0.009, $I^2 = 41\%$). In contrast, the patients carrying only one antibody were not significantly associated with smoking (Fig. 2A, 1 Ab *vs.* 0 Ab: $OR_{re} = 1.12$, $p_{re} = 0.4$, $I^2 = 56\%$). These associations were not different in the EAC and the prevalent RA cohorts (Supplementary Table S2).

Association of smoking with the presence of RF in the patients with one or two positive antibodies.
The patients carrying RF accounted for the smoking associations in the patients without the concurrent presence of the three antibodies. First, the specific RF association was found in RA patients with only one positive antibody (1 Ab) (Supplementary Table S3). In them, the patients carrying RF (RF^+_1) were associated with smoking (RF^+_1 vs. 0 Ab: OR = 1.28; 95% CI = 1.03–1.61; p = 0.03), whereas the RA patients carrying other

		Number of autoantibodies		
	0	**1**	**2**	**3**
IDIPAZ				
Non-smoker, n[b]	40	7	38	41
Smoker, n	14	14	40	49
PEARL				
Non-smoker, n	37	28	51	33
Smoker, n	22	17	36	40
IdISSC				
Non-smoker, n	90	60	75	51
Smoker, n	51	58	52	71
IDIS				
Non-smoker, n	83	81	138	72
Smoker, n	28	14	27	27
IDIVAL				
Non-smoker, n	88	39	41	38
Smoker, n	85	39	52	77
Rome				
Non-smoker, n	31	34	45	32
Smoker, n	31	33	53	50
Summary statistics				
OR	1 (ref.)	1.19	1.19	1.99
95% CI	—	0.91–1.55	0.94–1.52	1.56–2.54
p	—	0.21	0.15	2.48×10^{-08}
I^2, %	—	68.9	56.9	25.9
OR_{re}	1 (ref.)	1.22	1.22	1.99
95% CI_{re}	—	0.74–2.01	0.84–1.77	1.49–2.64
p_{re}	—	0.43	0.30	2.47×10^{-06}

Table 2. Association of smoking with seropositive patients with different numbers of autoantibodies in the new patient cohorts[a]. [a]The table presents the number of patients in each category in the upper part and the summary statistics obtained with meta-analysis in the lower part. The triple negative patients were used as reference (ref.) for the patients with 1, 2 or 3 antibodies. [b]n = number of subjects; OR = odds ratio; CI = confidence interval; I^2 = inconsistence; the re subscript indicates the random effects model.

positive antibodies (RF^-_1) were indistinguishable from the triple negative patients (RF^-_1 vs. 0 Ab: OR_{re} = 1.01; 95% CI = 0.65–1.56; p_{re} = 0.97). In addition, smoking was also associated with the presence of RF in the RA patients carrying two positive antibodies (2 Ab) (Supplementary Table S4). In effect, the patients in whom one of the two positive antibodies was RF (RF^+_2) were significantly associated with smoking (RF^+_2 vs. 0 Ab: OR = 1.30; 95% CI = 1.09–1.55; p = 0.004), whereas the other patients carrying two antibodies (RF^-_2) were undiscernible from the triple negative patients (RF^-_2 vs. 0 Ab: OR = 0.95; 95% CI = 0.64–1.39; p = 0.78). Similar analyses centered on the presence of ACPA or ACarPA did not show any significant association.

A notable finding of the preceding analysis was the very similar association (OR = 1.28 and 1.30) of smoking with patients carrying RF in the patients with only one positive antibody (RF^+_1) and with two positive antibodies (RF^+_2). This equivalence was confirmed when the associations of smoking with these two subgroups of patients were directly compared (RF^+_2 vs. RF^+_1: p = 0.9). As a consequence, the RF^+ patients carrying one or two positive antibodies were grouped (RF^+_{1+2}). As expected, smoking was associated with this unique RF^+ subgroup (Fig. 2A, RF^+_{1+2} vs. 0 Ab: p = 0.001, I^2 = 45%). An association that was slightly reinforced when all the RF^- patients were used as reference (Fig. 2B, RF^+_{1+2} vs. RF^-_{0+1+2}: p = 0.0002, I^2 = 0%). It should be remarked that the patients without RF in this latter analysis (RF^-_{0+1+2}) included patients positive for ACPA or ACarPA or both these antibodies. The previous associations were not significantly different in the EAC and the cohorts including prevalent RA patients (Supplementary Table S5).

Global exploratory analysis.
We used two exploratory techniques to check if a global analysis of all patient subgroups together was consistent with the sequential analyses in the preceding paragraphs without generating redundant statistical tests.

First, the relative frequencies of ever smokers and never smokers in each of the autoantibody-defined strata were displayed with a double-decker plot. Consistently with the results obtained in the sequential analyses, the smokers were enriched in the RF^+ patients relative to the corresponding RF^- patients in all the strata (Fig. 3).

The second exploratory analysis consisted of an exhaustive search of the best classification tree discriminating ever smokers from never smokers based on the antibodies and their combinations (Fig. 4). The first split of the tree was according to the concurrent presence of the 3 antibodies. The 1187 patients with concordant triple
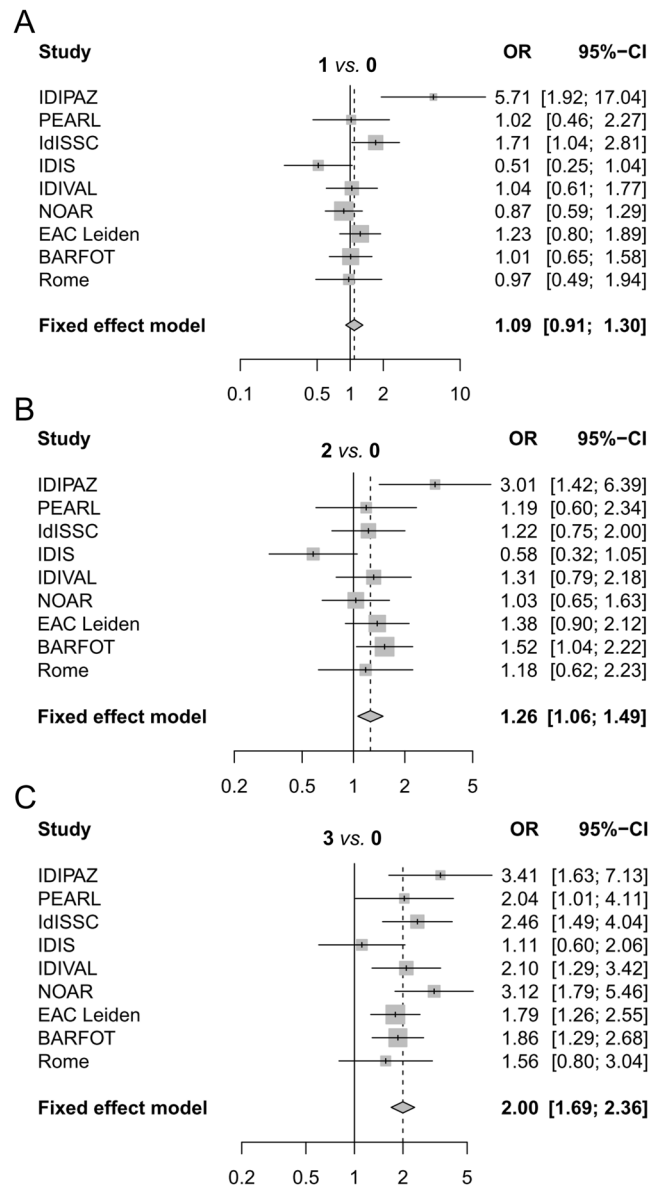
Figure 1. Smoking association with seropositivity for different numbers of autoantibodies in the combined meta-analysis. Forest plots showing the fixed-effect meta-analysis involving the RA patients that were positive for (**A**) one, (**B**) two, and (**C**) three autoantibodies. The patients without antibodies (0 Ab) were taken as reference. The OR corresponding to each cohort (Study) and its 95%-CI are provided and shown by a vertical tick and the length of a horizontal line, respectively. The area of the square around the tick is proportional to the size of the study. Summary OR and 95% CI are given in bold and represented as a vertical dashed line and a diamond, respectively. The OR scale is logarithmic.

seropositivity (3 Ab) contained 61% of smokers, whereas the remaining patients only 48% of smokers. No other antibody or combination contributed to the classification of the concordant patients. In contrast, two more divisions were observed in the non-concordant patients. The second split was on the presence of RF. The RF^+ group (RF^+_{1+2}) contained more smokers than the RF^- stratum (RF^-_{0+1+2}). The third split was unanticipated and complex. It divided the RF^+ patients into a smoker-enriched subgroup in whom RF was the only present antibody (RF^+_1), and a smoker-depleted subgroup in which RF was present concurrently with other antibodies (RF^+_2). This latter division reinforced the exclusivity for RF of the smoking association in the patients with one or two antibodies.

Discussion

Our main findings have been the predominant association of smoking with the concordant presence of the three RA antibodies, and the exclusive association with the presence of RF in the seropositive patients with one or two antibodies (RF^+_{1+2}). In addition, there was no significant association of smoking with the presence of ACPA or ACarPA. These findings have represented a significant advance. In van Wesemael et al.[20], the association of
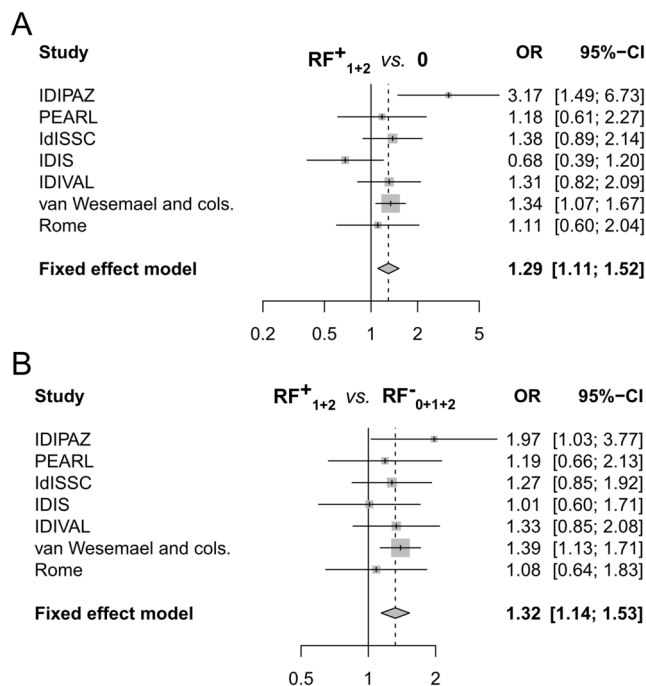
Figure 2. Association of smoking with the presence of RF in the RA patients carrying one or two antibodies. Forest plots showing the comparison of: (**A**) RF$^+$ patients that were positive for one or two antibodies (RF$^+_{1+2}$) with the patients without antibodies (0); and (**B**) the comparison of RF$^+$ patients that were positive for one or two antibodies (RF$^+_{1+2}$) with all the RF$^-$ patients (RF$^-_{0+1+2}$). No cohort-specific information was available for the three cohorts in van Wesemael *et al*. Plots follow Fig. 1 conventions.
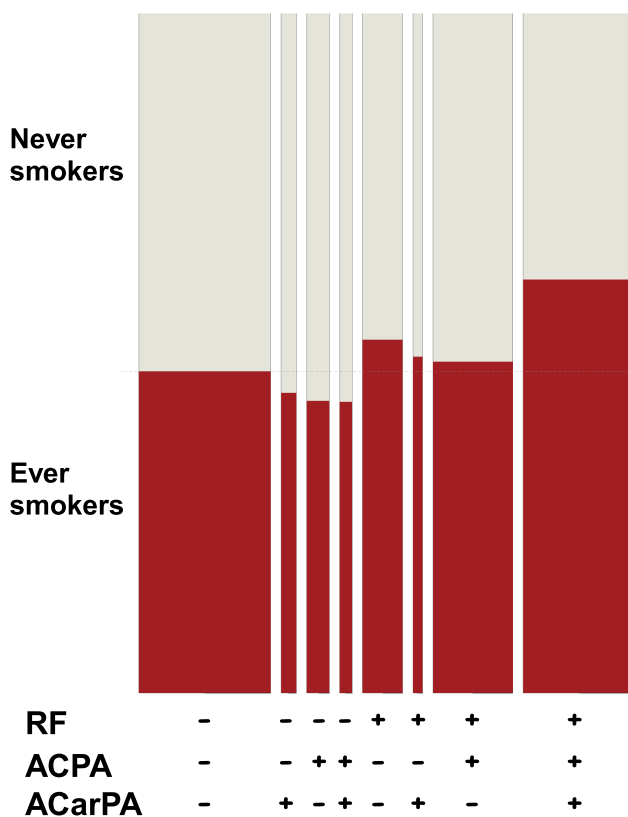
Figure 3. Smoking frequency in the patients stratified by the presence of the three autoantibodies. The double-decker plot divides the patients in rectangles proportional to the frequency in each subgroup. The width is proportional to the size of the antibody-defined patient subgroups, whereas the height of the red rectangles is proportional to the ever smokers within the antibody subgroups.
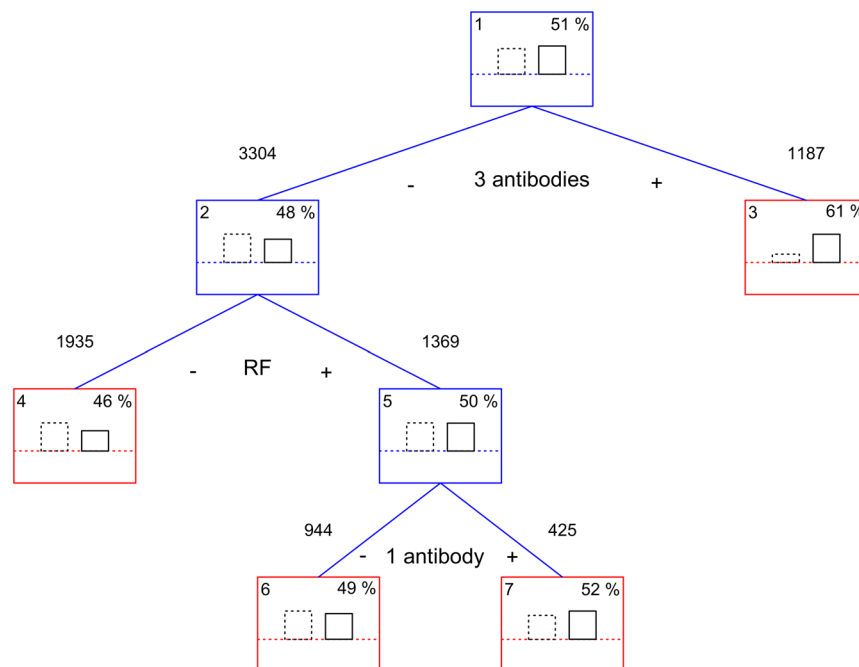
Figure 4. Classification tree of the patients by the presence of antibodies according to smoking. Each node shows its number (upper left corner), the percentage of ever smokers (upper right corner), and an histogram representing the frequency of ever smokers (continuous line) and never smokers (dotted line). Decision nodes are in blue whereas terminal nodes are in red. The splitting conditions are displayed below the node with indication of the value determining each side of the split and the number of subjects sent to the child node.

smoking with RA could represent a gradual increase in the strength of association with the number of antibodies. The evidence presented here excludes this putative mechanism by the demonstration of the exclusive association of smoking with RF independently of the presence or absence of other antibodies in the patients carrying one or two antibodies, and by the clear distinction between the patients with three antibodies as a separate classification to other seropositive patients. This fundamental insight determines the nature of the models aiming to explain the effect of smoking on RA susceptibility. In addition, our analysis has made more understandable the relationship between smoking and the antibodies thanks to the exploratory techniques. They also showed that the results of the statistical tests were faithful to the data.

The originality of our findings is reflected in the absence of any other study analyzing the association of smoking with combinations of the three antibodies included here and in van Wesemael et al.[20]. Therefore, we considered the studies assessing the combination of RF and ACPA as antecedents. We have found only reports from four large cohorts[20,33–35], one of them included in van Wesemael et al. but different from the RA cohorts included in our meta-analysis. The four cohorts were large, with ≈9500 healthy Japanese subjects[20], ≈2000 UK RA patients[33], ≈1500 USA RA patients[34], and ≈3600 Swedish RA patients[35]. The three first have shown a significant association of smoking only with the concurrent presence of the two antibodies, not with any of them in isolation. The forth by Hedstrom et al. showed a stronger association with the concurrent presence of RF and ACPA, followed by RF and less significantly by ACPA (more on the results of this study below). These results and those of another study with two smaller UK RA collections[36] are fully compatible with our findings. In addition, the bibliographic search brought to our attention another important fact. None of the studies that support the pathogenic model linking smoking with RA through the production of ACPA has accounted for the association of smoking with the concurrent presence of ACPA and RF[10,11,37–39] until very recently[35].

We do not know the mechanism behind the predominant association of smoking with the concurrent presence of RA autoantibodies. However, it is well-known that the status of the RA autoantibodies is much more concordant than at random[25,28,40]. This circumstance reveals the existence of pathogenic mechanisms that are shared by the various antibodies. Some of these mechanisms contribute to epitope spreading, which characterizes the progression of T and B cell responses in autoimmune diseases including the preclinical phase of RA[3–5,41,42]. In effect, the earliest seropositive samples from patients that will develop RA years later often recognize a single epitope, whereas samples taken near the clinical onset recognize multiple epitopes[3–5]. Therefore, smoking could promote concordant seropositivity by broadening and accelerating epitope spreading.

The known factors affecting epitope spreading include the availability of epitope-specific lymphocytes, reflecting incompetent tolerance, and favorable T - B cell interactions and antigen presentation[41,42]. The latter interactions could be boosted by bystander activation, tissue damage and inflammation. Therefore, smoking could promote epitope spreading through these multiple mechanisms. This is possible because the triggering of inflammation and tissue damage, the recruiting and activation of neutrophils, monocytes, and macrophages, and abnormalities in NK, dendritic cells, B cells and many subtypes of T lymphocytes are some of the many effects of

smoking on the immune responses[43,44]. The overall balance of this range of actions is an increased predisposition to autoimmunity and the production of autoantibodies[43,44]. Specifically, smoking has been associated with the production of anti-dsDNA in SLE[45], anti-Jo1 in inflammatory myopathies[46] and of RF and other autoantibodies in smokers without any autoimmune disease[20,47–50].

The sharing of immunological mechanisms between the autoantibodies is also the most likely explanation for the correlation between their antibody titers, a correlation that we have also observed in our analyses[25,51–53]. These correlations have been observed independently of the disease stage and, most significantly, to be maintained as parallel titer decreases in response to treatment revealing that they respond similarly to the control of inflammation[51–53]. These correlations lack any specific direction and extend to the thresholds for establishing the positive/negative status (Supplementary Table S6). Therefore, they are unlikely to denote any hierarchy of precedence or causality between the antibodies. Just recently, some of these examples of shared and overlapping immunological mechanisms have been characterized as the convergent pathways model of RA pathogenesis[54].

The association of smoking with RF in the patients with one of two antibodies suggests that RF could precede other autoantibodies in the smokers that will become RA patients. However, we lack support for this interpretation given the cross-sectional nature of our sample collections. In addition, the studies of preclinical RA samples have been discordant in the order of antibody appearance: IgA RF and IgM RF were the first antibodies in a Swedish cohort[55], whereas ACPA[4] and ACarPA[56] preceded IgM RF in a Dutch cohort, and again IgG ACPA preceded RF (no ACarPA analysis included) in American military[57]. Similarly, the presence of RF in smokers without RA that has been known for more than two decades[49,50], cannot be taken as evidence of RF preceding the other autoantibodies because recent studies have found also an increased presence of ACPA in smokers without RA[20,58]. In consequence, we will need to wait for new studies to solve this question.

Independently of the order of appearance, we need an explanation for the specific association of smoking with RF. A couple of possible mechanisms have already been proposed. One of them was originally developed to explain the production of autoantibodies in smokers without autoimmune diseases[48]. It starts by the induction of heat-shock protein 70 (Hsp70) expression and of antibodies against Hsp70 by smoking. In the next step, the production of RF is triggered by the two signals given by the Hsp70 immune complexes (IC): through the BCR recognizing the anti-Hsp70 IgG, and through CD91 binding Hsp70. These details were delineated in mouse experiments[48], but their reproducibility in RA patients is unclear. A second hypothesis proposes that smoking leads to increased lung production of IgG, which would be recognized by RF in its native form and by ACPA and ACarPA as the citrullinated and carbamylated modified IgGH fragment, respectively[59]. This hypothesis has the appeal of the simplicity of considering a single protein as the link between the different RA antibodies. Accordingly, it has been characterized as the common antigen model[54]. However, only RF is known to recognize the IgGH fragment, in the form of IgGH/HLA class II complexes[60], whereas the binding of the modified IgGH fragment by ACPA or ACarPA has not been reported. The two hypotheses could become the starting point for future experiments.

The association of smoking with RF positive RA was known before the association with the ACPA positive patients[6,7], however, the latter displaced RF from the focus of attention[1,2]. Probably, the peculiar nature of RF has some role in this displacement. In effect, RF can be described as an antibody against IC with a role in the development of the early antibody repertoire that in the adult can be induced in the course of sustained immunological responses that include other diseases besides RA[61–63]. However, the nature of the RF in patients with RA and non-RA subjects have significant differences. In RA, the RF production is sustained and reach higher titers; also, the range of Ig V genes that are used is broader, and the response shows signs of maturation as isotype switch and mutations changing the sequence of the CDRs, characteristics that are absent or restricted in the RF of healthy subjects[62,63]. It is understood that ACPA and citrullinated proteins, or ACarPA and carbamylated proteins, are the IC recognized in RA, but other antibodies are possible as the anti-Hsp70 antibodies (in relation to smoking) and microbial antigens (including those from viruses and the mucosal microbiota). These diverse IC could act as disease triggers[48,61]. Once RF binds to the IC, the IC become larger and capable of more efficient immune stimulation. A model of RA in which two waves of IC formation, without RF and with RF, followed by complement activation and the production of inflammatory and chemotactic mediators has been proposed[61]. It is also possible that the RF-specific B cells have a critical role in the early phases of the autoimmune response before high titers of RF have been produced[63]. In effect, the RF specific B cells are abundant in healthy subjects and the only B cells capable of efficient presentation to T cells of the antigens trapped in IC[64]. These various roles of RF and the RF-specific B cells could contribute to epitope spreading and the concordance and correlation between the RA autoantibodies.

The fact that our meta-analysis did not detect any specific association of smoking with the presence of ACPA in the patients without RF (OR = 0.95, 95% CI = 0.78–1.20) does not exclude association with a subset of the ACPA$^+$ patients. Examples of such subsets could be the patients with shared epitope HLA alleles or patients with heavy and current smoking. Concerning the HLA, the presence of the shared epitope is specifically associated with the presence of ACPA and there are arguments to think it could potentiate the smoking association. The most clear argument comes from a recent study by Hedstrom et al. showing association with ACPA$^+$/RF$^-$ [35], whereas previous studies failed to stratify by the two antibodies or did not find a significant association[10,11,33,34,36–39]. The other example is supported by the same recent study, which showed differential association of heavy smokers and light smokers with RA. A difference that was more marked in the ACPA$^+$/RF$^+$ patients, followed by the ACPA$^-$/RF$^+$ patients and finally the ACPA$^+$/RF$^-$ patients[35]. This differential association is in agreement with our results placing the concordant seropositivity at the top and RF afterwards. However, the ACPA$^+$/RF$^-$ patients in Hedstrom et al. were significantly associated with smoking and they were not in our meta-analysis.

Our study lacks healthy controls and therefore, we were restricted to comparisons between patient subsets. However, this is a minor limitation because the association of smoking with seropositive RA patients is well-established[6,7]. In addition, the study lacks information on smoking intensity and HLA alleles. These two

types of information could have provided additional insight into the relationship between smoking and the different autoantibodies. In any case, a whole analysis of the potential interactions between smoking and the HLA in the three autoantibodies is not yet possible because only the HLA alleles associated with ACPA are well-defined[65]. The other antibodies, ACarPA and RF, could be specifically associated with HLA-DRB1 alleles that are not included in the shared epitope, but that have not been yet completely defined[66–68]. Finally, we should signal that the classification tree in Fig. 4 is not appropriate to separate smokers and non-smokers. This algorithm was only intended as a tool to explore the relationships between smoking and the antibodies present in the data.

In summary, we can conclude that smoking is predominantly and reproducibly associated with the triple, RF, ACPA, and ACarPA, concordant seropositive RA. A result that highlights the need to consider the mechanisms leading to concurrent seropositivity. In the patients that are not concordant for the three antibodies, smoking was exclusively associated with RF positivity in our meta-analysis. This latter association is weaker than the association with the triple concordant patients and its exclusivity needs to be replicated, ideally in studies counting with detailed information on smoking intensity. These results call for a pathogenic model that incorporates the predominant association with multiple antibodies, which could be explained by accelerated and broadened epitope spreading reflecting some of the actions of smoking on the immune system. In addition, an explanation for the particular association of smoking with the production of RF should be sought.

References

1. Derksen, V., Huizinga, T. W. J. & van der Woude, D. The role of autoantibodies in the pathophysiology of rheumatoid arthritis. *Semin. Immunopathol.* **39**, 437–446, https://doi.org/10.1007/s00281-017-0627-z (2017).
2. Malmstrom, V., Catrina, A. I. & Klareskog, L. The immunopathogenesis of seropositive rheumatoid arthritis: from triggering to targeting. *Nat. Rev. Immunol.* **17**, 60–75, https://doi.org/10.1038/nri.2016.124 (2017).
3. van der Woude, D. *et al*. Epitope spreading of the anti-citrullinated protein antibody response occurs before disease onset and is associated with the disease course of early arthritis. *Ann. Rheum. Dis.* **69**, 1554–1561, https://doi.org/10.1136/ard.2009.124537 (2010).
4. van de Stadt, L. A. *et al*. Development of the anti-citrullinated protein antibody repertoire prior to the onset of rheumatoid arthritis. *Arthritis Rheum.* **63**, 3226–3233, https://doi.org/10.1002/art.30537 (2011).
5. Sokolove, J. *et al*. Autoantibody epitope spreading in the pre-clinical phase predicts progression to rheumatoid arthritis. *Plos One* **7**, e35296, https://doi.org/10.1371/journal.pone.0035296 (2012).
6. Sugiyama, D. *et al*. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. *Ann. Rheum. Dis.* **69**, 70–81, https://doi.org/10.1136/ard.2008.096487 (2010).
7. Di Giuseppe, D., Discacciati, A., Orsini, N. & Wolk, A. Cigarette smoking and risk of rheumatoid arthritis: a dose-response meta-analysis. *Arthritis Res. Ther.* **16**, R61, https://doi.org/10.1186/ar4498 (2014).
8. Padyukov, L., Silva, C., Stolt, P., Alfredsson, L. & Klareskog, L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum.* **50**, 3085–3092, https://doi.org/10.1002/art.20553 (2004).
9. Criswell, L. A. *et al*. Smoking interacts with genetic risk factors in the development of rheumatoid arthritis among older Caucasian women. *Ann. Rheum. Dis.* **65**, 1163–1167, https://doi.org/10.1136/ard.2005.049676 (2006).
10. Kallberg, H. *et al*. Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. *Am. J. Hum. Genet.* **80**, 867–875, https://doi.org/10.1086/516736 (2007).
11. Klareskog, L. *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.* **54**, 38–46, https://doi.org/10.1002/art.21575 (2006).
12. Makrygiannakis, D. *et al*. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Ann. Rheum. Dis.* **67**, 1488–1492, https://doi.org/10.1136/ard.2007.075192 (2008).
13. Damgaard, D. *et al*. Smoking is associated with increased levels of extracellular peptidylarginine deiminase 2 (PAD2) in the lungs. *Clin. Exp. Rheumatol.* **33**, 405–408 (2015).
14. Bongartz, T. *et al*. Citrullination in extra-articular manifestations of rheumatoid arthritis. *Rheumatol.* **46**, 70–75, https://doi.org/10.1093/rheumatology/kel202 (2007).
15. Lugli, E. B. *et al*. Expression of citrulline and homocitrulline residues in the lungs of non-smokers and smokers: implications for autoimmunity in rheumatoid arthritis. *Arthritis Res. Ther.* **17**, 9, https://doi.org/10.1186/s13075-015-0520-x (2015).
16. Demoruelle, M. K. *et al*. Brief report: airways abnormalities and rheumatoid arthritis-related autoantibodies in subjects without arthritis: early injury or initiating site of autoimmunity? *Arthritis Rheum.* **64**, 1756–1761, https://doi.org/10.1002/art.34344 (2012).
17. Reynisdottir, G. *et al*. Structural changes and antibody enrichment in the lungs are early features of anti-citrullinated protein antibody-positive rheumatoid arthritis. *Arthritis Rheumatol.* **66**, 31–39, https://doi.org/10.1002/art.38201 (2014).
18. Reynisdottir, G. *et al*. Signs of immune activation and local inflammation are present in the bronchial tissue of patients with untreated early rheumatoid arthritis. *Ann. Rheum. Dis.* **75**, 1722–1727, https://doi.org/10.1136/annrheumdis-2015-208216 (2016).
19. Willis, V. C. *et al*. Sputum autoantibodies in patients with established rheumatoid arthritis and subjects at risk of future clinically apparent disease. *Arthritis Rheum.* **65**, 2545–2554, https://doi.org/10.1002/art.38066 (2013).
20. van Wesemael, T. J. *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.* **18**, 285, https://doi.org/10.1186/s13075-016-1177-9 (2016).
21. Pecani, A. *et al*. Prevalence, sensitivity and specificity of antibodies against carbamylated proteins in a monocentric cohort of patients with rheumatoid arthritis and other autoimmune rheumatic diseases. *Arthritis Res. Ther.* **18**, 276, https://doi.org/10.1186/s13075-016-1173-0 (2016).
22. Orozco, G. *et al*. Auto-antibodies, HLA and PTPN22: susceptibility markers for rheumatoid arthritis. *Rheumatol.* **47**, 138–141, https://doi.org/10.1093/rheumatology/kem343 (2008).
23. Gonzalez-Alvaro, I. *et al*. Interleukin 15 levels in serum may predict a severe disease course in patients with early arthritis. *Plos One* **6**, e29492, https://doi.org/10.1371/journal.pone.0029492 (2011).
24. Arnett, F. C. *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324, https://doi.org/10.1002/art.1780310302 (1988).
25. Montes, A. *et al*. Anti-Carbamylated Protein Antibodies as a Reproducible Independent Type of Rheumatoid Arthritis Autoantibodies. *Plos One* **11**, e0161141, https://doi.org/10.1371/journal.pone.0161141 (2016).
26. Abasolo, L. *et al*. Influence of demographic and clinical factors on the mortality rate of a rheumatoid arthritis cohort: A 20-year survival study. *Semin. Arthritis Rheum.* **45**, 533–538, https://doi.org/10.1016/j.semarthrit.2015.10.016 (2016).

27. Corrales, A. *et al*. Cardiovascular risk stratification in rheumatic diseases: carotid ultrasound is more sensitive than Coronary Artery Calcification Score to detect subclinical atherosclerosis in patients with rheumatoid arthritis. *Ann. Rheum. Dis.* **72**, 1764–1770, https://doi.org/10.1136/annrheumdis-2013-203688 (2013).

28. Shi, J. *et al*. Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. *Proc. Natl Acad. Sci. USA* **108**, 17372–17377, https://doi.org/10.1073/pnas.1114465108 (2011).

29. Schwarzer, G. meta: An R package for meta-analysis. *R News* **7**, 40-45 (2007). https://cran.r-project.org/doc/Rnews/Rnews_2007-3.pdf.

30. Review Manager (RevMan) v. Version 5.3 (The Cochrane Collaboration, Copenhagen: The Nordic Cochrane Centre, (2014).

31. Hofmann, H. Generalized Odds Ratios for Visual Modeling. *J. Computational Graph. Stat.* **10**, 628–640, https://doi.org/10.1198/106186001317243368 (2001).

32. Montes, A. *et al*. Particular association of clinical and genetic features with autoimmunity to citrullinated alpha-enolase in rheumatoid arthritis. *Arthritis Rheum.* **63**, 654–661, https://doi.org/10.1002/art.30186 (2011).

33. Morgan, A. W. *et al*. Reevaluation of the interaction between HLA-DRB1 shared epitope alleles, PTPN22, and smoking in determining susceptibility to autoantibody-positive and autoantibody-negative rheumatoid arthritis in a large UK Caucasian population. *Arthritis Rheum.* **60**, 2565–2576, https://doi.org/10.1002/art.24752 (2009).

34. Sokolove, J. *et al*. Rheumatoid factor as a potentiator of anti-citrullinated protein antibody-mediated inflammation in rheumatoid arthritis. *Arthritis Rheumatol.* **66**, 813–821, https://doi.org/10.1002/art.38307 (2014).

35. Hedstrom, A. K., 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.* **71**, 1504–1511, https://doi.org/10.1002/art.40852 (2019).

36. 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* **12**, e0180655, https://doi.org/10.1371/journal.pone.0180655 (2017).

37. Linn-Rasker, S. P. *et al*. Smoking is a risk factor for anti-CCP antibodies only in rheumatoid arthritis patients who carry HLA-DRB1 shared epitope alleles. *Ann. Rheum. Dis.* **65**, 366–371, https://doi.org/10.1136/ard.2005.041079 (2006).

38. Pedersen, M. *et al*. Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control study in Denmark. *Arthritis Rheum.* **56**, 1446–1453, https://doi.org/10.1002/art.22597 (2007).

39. Lee, H. S. *et al*. Interaction between smoking, the shared epitope, and anti-cyclic citrullinated peptide: a mixed picture in three large North American rheumatoid arthritis cohorts. *Arthritis Rheum.* **56**, 1745–1753, https://doi.org/10.1002/art.22703 (2007).

40. Verheul, M. K. *et al*. Triple Positivity for Anti-Citrullinated Protein Autoantibodies, Rheumatoid Factor, and Anti-Carbamylated Protein Antibodies Conferring High Specificity for Rheumatoid Arthritis: Implications for Very Early Identification of At-Risk Individuals. *Arthritis Rheumatol.* **70**, 1721–1731, https://doi.org/10.1002/art.40562 (2018).

41. Vanderlugt, C. L. & Miller, S. D. Epitope spreading in immune-mediated diseases: implications for immunotherapy. *Nat. Rev. Immunol.* **2**, 85–95, https://doi.org/10.1038/nri724 (2002).

42. Cornaby, C. *et al*. B cell epitope spreading: mechanisms and contribution to autoimmune diseases. *Immunol. Lett.* **163**, 56–68, https://doi.org/10.1016/j.imlet.2014.11.001 (2015).

43. Costenbader, K. H. & Karlson, E. W. Cigarette smoking and autoimmune disease: what can we learn from epidemiology? *Lupus* **15**, 737–745, https://doi.org/10.1177/0961203306069344 (2006).

44. Qiu, F. *et al*. Impacts of cigarette smoking on immune responsiveness: Up and down or upside down? *Oncotarget* **8**, 268–284, https://doi.org/10.18632/oncotarget.13613 (2017).

45. Freemer, M. M., King, T. E. Jr. & Criswell, L. A. Association of smoking with dsDNA autoantibody production in systemic lupus erythematosus. *Ann. Rheum. Dis.* **65**, 581–584, https://doi.org/10.1136/ard.2005.039438 (2006).

46. Chinoy, H. *et al*. Interaction of HLA-DRB1*03 and smoking for the development of anti-Jo-1 antibodies in adult idiopathic inflammatory myopathies: a European-wide case study. *Ann. Rheum. Dis.* **71**, 961–965, https://doi.org/10.1136/annrheumdis-2011-200182 (2012).

47. Korpilahde, T. *et al*. Smoking history and serum cotinine and thiocyanate concentrations as determinants of rheumatoid factor in non-rheumatoid subjects. *Rheumatol.* **43**, 1424–1428, https://doi.org/10.1093/rheumatology/keh365 (2004).

48. Newkirk, M. M. *et al*. Chronic smoke exposure induces rheumatoid factor and anti-heat shock protein 70 autoantibodies in susceptible mice and humans with lung disease. *Eur. J. Immunol.* **42**, 1051–1061, https://doi.org/10.1002/eji.201141856 (2012).

49. Jonsson, T., Thorsteinsson, J. & Valdimarsson, H. Does smoking stimulate rheumatoid factor production in non-rheumatic individuals? *APMIS* **106**, 970–974, https://doi.org/10.1111/j.1699-0463.1998.tb00247.x (1998).

50. Tuomi, T., Heliovaara, M., Palosuo, T. & Aho, K. Smoking, lung function, and rheumatoid factors. *Ann. Rheum. Dis.* **49**, 753–756, https://doi.org/10.1136/ard.49.10.753 (1990).

51. Alessandri, C. *et al*. Decrease of anti-cyclic citrullinated peptide antibodies and rheumatoid factor following anti-TNFalpha therapy (infliximab) in rheumatoid arthritis is associated with clinical improvement. *Ann. Rheum. Dis.* **63**, 1218–1221, https://doi.org/10.1136/ard.2003.014647 (2004).

52. Bobbio-Pallavicini, F. *et al*. Autoantibody profile in rheumatoid arthritis during long-term infliximab treatment. *Arthritis Res. Ther.* **6**, R264–272, https://doi.org/10.1186/ar1173 (2004).

53. Atzeni, F. *et al*. Adalimumab clinical efficacy is associated with rheumatoid factor and anti-cyclic citrullinated peptide antibody titer reduction: a one-year prospective study. *Arthritis Res. Ther.* **8**, R3, https://doi.org/10.1186/ar1851 (2006).

54. Shelef, M. A. New Relationships for Old Autoantibodies in Rheumatoid Arthritis. *Arthritis Rheumatol.* **71**, 1396–1399, https://doi.org/10.1002/art.40879 (2019).

55. Brink, M. *et al*. Rheumatoid factor isotypes in relation to antibodies against citrullinated peptides and carbamylated proteins before the onset of rheumatoid arthritis. *Arthritis Res. Ther.* **18**, 43, https://doi.org/10.1186/s13075-016-0940-2 (2016).

56. Shi, J. *et al*. Anti-carbamylated protein (anti-CarP) antibodies precede the onset of rheumatoid arthritis. *Ann. Rheum. Dis.* **73**, 780–783, https://doi.org/10.1136/annrheumdis-2013-204154 (2014).

57. Kelmenson, L. B. *et al*. Timing of elevations of autoantibody isotypes in rheumatoid arthritis prior to disease diagnosis. *Arthritis Rheumatol* **0**, https://doi.org/10.1002/art.41091 (2019).

58. van Zanten, A. *et al*. Presence of anticitrullinated protein antibodies in a large population-based cohort from the Netherlands. *Ann. Rheum. Dis.* **76**, 1184–1190, https://doi.org/10.1136/annrheumdis-2016-209991 (2017).

59. Hutchinson, D., Murphy, D. & Clarke, A. & Eggleton, P. Are Rheumatoid Factor, Anti-Citrullinated Protein Antibodies, and Anti-Carbamylated Protein Antibodies Linked by Posttranslational Modification of IgG? Comment on the Article by Koppejan et al. *Arthritis Rheumatol.* **68**, 2825–2826, https://doi.org/10.1002/art.39832 (2016).

60. Jin, H. *et al*. Autoantibodies to IgG/HLA class II complexes are associated with rheumatoid arthritis susceptibility. *Proc. Natl Acad. Sci. USA* **111**, 3787–3792, https://doi.org/10.1073/pnas.1401105111 (2014).

61. Tan, E. M. & Smolen, J. S. Historical observations contributing insights on etiopathogenesis of rheumatoid arthritis and role of rheumatoid factor. *J. Exp. Med.* **213**, 1937–1950, https://doi.org/10.1084/jem.20160792 (2016).

62. Mageed, R. A., Borretzen, M., Moyes, S. P., Thompson, K. M. & Natvig, J. B. Rheumatoid factor autoantibodies in health and disease. *Ann. N. Y. Acad. Sci.* **815**, 296–311, https://doi.org/10.1111/j.1749-6632.1997.tb52071.x (1997).

63. Carson, D. A., Chen, P. P. & Kipps, T. J. New roles for rheumatoid factor. *J. Clin. Invest.* **87**, 379–383, https://doi.org/10.1172/JCI115007 (1991).

64. Roosnek, E. & Lanzavecchia, A. Efficient and selective presentation of antigen-antibody complexes by rheumatoid factor B cells. *J. Exp. Med.* **173**, 487–489, https://doi.org/10.1084/jem.173.2.487 (1991).
65. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44**, 291–296, https://doi.org/10.1038/ng.1076 (2012).
66. Regueiro, C. *et al.* Specific Association of HLA-DRB1*03 With Anti-Carbamylated Protein Antibodies in Patients With Rheumatoid Arthritis. *Arthritis Rheumatol.* **71**, 331–339, https://doi.org/10.1002/art.40738 (2019).
67. Jiang, X. *et al.* Anti-CarP antibodies in two large cohorts of patients with rheumatoid arthritis and their relationship to genetic risk factors, cigarette smoking and other autoantibodies. *Ann. Rheum. Dis.* **73**, 1761–1768, https://doi.org/10.1136/annrheumdis-2013-205109 (2014).
68. Han, B. *et al.* Fine Mapping Seronegative and Seropositive Rheumatoid Arthritis to Shared and Distinct HLA Alleles by Adjusting for the Effects of Heterogeneity. *Am. J. Hum. Genet.* **94**, 522–532, https://doi.org/10.1016/j.ajhg.2014.02.013 (2014).

Acknowledgements

Author contributions

C.R., A.B., I.G.-A. and A.G. designed the study. C.R. performed laboratory determinations; L.R.-R., R.L.-M., L.N., A.T.-M., E.P.-P., A.C., A.V., Y.L.-G., L.A., S.R.-M., A.M.O., E.H., A.M.-F., C.C., A.M.-V., A.B., I.G.-A., M.A.G.-G. and B.F.-G. provided patients and clinical and laboratory data. C.R. and A.G. performed analysis of the results and wrote the first draft of the manuscript. L.R.-R., R.L.-M., L.N., A.T.-M., E.P.-P., A.C., A.V., Y.L.-G., L.A., S.R.-M., A.M.O., E.H., A.M.-F., C.C., A.M.-V., A.B., I.G.-A., M.A.G.-G. and B.F.-G. revised the manuscript for important content. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-60305-x.

Correspondence and requests for materials should be addressed to A.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.