



**Universiteit  
Leiden**  
The Netherlands

## **Genetic studies in rheumatoid arthritis**

Kurreeman, F.

### **Citation**

Kurreeman, F. (2009, November 4). *Genetic studies in rheumatoid arthritis*.  
Retrieved from <https://hdl.handle.net/1887/14323>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/14323>

**Note:** To cite this publication please use the final published version (if applicable).

## **Chapter 2**

### **A Candidate Gene Approach Identifies the TRAF1/C5 Region as a Risk Factor for Rheumatoid Arthritis**

F.A.S. Kurreeman

L. Padyukov

R.B. Marques

S.J. Schrodi

M. Seddighzadeh

G. Stoeken-Rijsbergen

A.H.M. van der Helm-van Mil

C.F. Allaart

W. Verduyn

J. Houwing-Duistermaat

L. Alfredsson

A.B. Begovich

L. Klareskog

T.W.J. Huizinga

R.E.M. Toes

Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands  
Rheumatology Unit, Department of Medicine, Karolinska Institute at Karolinska Hospital, Stockholm, Sweden; Celera, Alameda, California, United States of America  
Department of Immunohaematology and Bloodbank, Leiden University Medical Center, Leiden, The Netherlands  
Institute of Environmental Medicine, Department of Medical Statistics, Leiden University Medical Centre, Leiden, The Netherlands

## Abstract

### Background

Rheumatoid arthritis (RA) is a chronic autoimmune disorder affecting ~1% of the population. The disease results from the interplay between an individual's genetic background and unknown environmental triggers. Although human leukocyte antigens (HLAs) account for ~30% of the heritable risk, the identities of non-HLA genes explaining the remainder of the genetic component are largely unknown. Based on functional data in mice, we hypothesized that the immune-related genes *complement component 5 (C5)* and/or *TNF receptor-associated factor 1 (TRAF1)*, located on Chromosome 9q33-34, would represent relevant candidate genes for RA. We therefore aimed to investigate whether this locus would play a role in RA.

### Methods and Findings

We performed a multitiered case-control study using 40 single-nucleotide polymorphisms (SNPs) from the *TRAF1* and *C5 (TRAF1/C5)* region in a set of 290 RA patients and 254 unaffected participants (controls) of Dutch origin. Stepwise replication of significant SNPs was performed in three independent sample sets from the Netherlands ( $n_{\text{cases/controls}} = 454/270$ ), Sweden ( $n_{\text{cases/controls}} = 1,500/1,000$ ) and US ( $n_{\text{cases/controls}} = 475/475$ ). We observed a significant association ( $p < 0.05$ ) of SNPs located in a haplotype block that encompasses a 65 kb region including the 3' end of *C5* as well as *TRAF1*. A sliding window analysis revealed an association peak at an intergenic region located ~10 kb from both *C5* and *TRAF1*. This peak, defined by SNP14/rs10818488, was confirmed in a total of 2,719 RA patients and 1,999 controls (odds ratio<sub>common</sub> = 1.28, 95% confidence interval 1.17–1.39,  $p_{\text{combined}} = 1.40 \times 10^{-8}$ ) with a population attributable risk of 6.1%. The A (minor susceptibility) allele of this SNP also significantly correlates with increased disease progression as determined by radiographic damage over time in RA patients ( $p = 0.008$ ).

### Conclusions

Using a candidate-gene approach we have identified a novel genetic risk factor for RA. Our findings indicate that a polymorphism in the *TRAF1/C5* region increases the susceptibility to and severity of RA, possibly by influencing the structure, function, and/or expression levels of *TRAF1* and/or *C5*.

## Introduction

Rheumatoid arthritis (RA) is characterized by chronic inflammation and destruction of the synovial joints leading to progressive joint damage and disability. The disease has a complex etiology, including a wide spectrum of clinical manifestations, variability in disease severity and/or progression, and differential response to a range of therapies. This heterogeneous phenotype suggests the involvement of both environmental and genetic factors [1], where the genetic component of RA has been estimated to be between 50%–60% [2,3]. Identification of disease-associated genes is important as it will guide our understanding of the biological pathways underlying polygenic diseases and the development of potential novel therapeutic targets.

The most prominent genetic association in RA is confined to the *human leukocyte antigen (HLA)* locus. Although this association has been known for almost 30 years, and although the underlying mechanism is still not understood, it has been replicated in multiple studies [2,4]. The identification of RA-associated genes outside of the *HLA* region, however, has been a challenge. Recently one such gene, *protein tyrosine phosphatase, non-receptor type 22 (lymphoid) (PTPN22)*, was identified in the first step of a large genetic-association study utilizing putative functional SNPs [5]. The gene product encoded by *PTPN22* is, like the *HLA* locus, involved in T cell-mediated immune responses. However, other immune components are also thought to play a pivotal role in RA, as demonstrated by the beneficial effects of treatment with agents that block proinflammatory cytokines, such as tumor necrosis factor  $\alpha$  (TNF $\alpha$ ) [6]. Moreover, in several experimental animal models for RA, innate immune responses mediated by a diversity of players have been implicated in arthritis. In this respect, a prominent role for the complement system has been identified as mice deficient in complement factors are resistant to arthritis, and as it has been shown that targeting *complement component 5 (C5)* by antibodies prevents the onset of arthritis and reduces the clinical severity in mouse models for arthritis [7,8]. Likewise, the observation that high levels of C5a, a potent chemoattractant, are found in synovial fluid of RA patients combined with the fact that *C5a receptor*-deficient mice are also resistant to arthritis induction, indicate a central role for these mediators in arthritis [9,10]. A genome scan of mice that were or were not susceptible to antibody-induced arthritis revealed that the main genetic influence detected in this model maps to the *C5* region [11].

These functional data in mice inspired us to hypothesize that the *C5* region would be a contributing factor in RA. Therefore, we searched for further evidence by first addressing the question of whether any genetic indications exist that implicate the involvement of this region in RA. A conventional linkage study using microsatellite markers identified a linkage peak in the vicinity of the *C5* region [12]. Although this study provided weak evidence for linkage (logarithm of the odds score, LOD 1.8), it further boosted our interest in this region. *C5* is located next to *TNF receptor-associated factor 1 (TRAF1)*, an essential effector of the TNF signaling cascade. Since TNF blockade represents a powerful intervention in both mice and humans for the treatment of arthritis, it provided an additional rationale to explore this genetic region encoding *C5* and *TRAF1*, which are adjacent to each other on Chromosome 9q33-34. We therefore sought to investigate whether these candidate genes, which are important immune mediators, would play a role in RA.

## Methods

### *Study Populations*

All RA patients in all sets in this study met the American College of Rheumatology 1987 revised criteria for RA [13]. Sample set 1 cases consisted of 290 RA patients consecutively included from the out-patient clinic of the Leiden University Hospital in 1994 [14] and 254 controls randomly selected by the section Immunogenetics and Transplantation Immunology of Leiden University Medical Center, Leiden, The Netherlands (ITI).

Since 89% of the first set of RA patients were rheumatoid factor (RF) positive, we genotyped an independent sample set 2, which consisted of 454 RF-positive patients from two inception cohorts of early arthritis patients (EAC and BEST) and a second set of 270 randomly selected Dutch blood donors from ITI. Briefly, the EAC consists of patients included from 1993 onwards and originating from a health care region of about 400,000 inhabitants in the western part of The Netherlands. General practitioners were encouraged to refer patients directly when arthritis was suspected. Patients were included when the symptom duration was less than 2 y. Patients from the BEST study were recruited between March 2000 and August 2002 at 20 centers in the western part of The Netherlands. Patients had a maximum disease duration of 2 y, were at least 18 y of age, and had active disease (defined as  $\geq$  six swollen joints,  $\geq$  six tender joints, and either an erythrocyte sedimentation rate of  $\geq$  28 mm/h or a global health assessment score of  $\geq$  20 on a 100-mm visual analog scale, where 0 = best and 100 = worst). Only patients with a diagnosis of RA were included in the present study. These cohorts are further described in detail in other reports [15,16]. Sample set 3 consisted of 1,500 RA patients (70% RF-positive) and 1,000 unaffected participants (controls) from the Swedish EIRA study as previously described [17]. Briefly, RA patients and controls aged 18–70 y during May 1996 to December 2003 from a geographically defined area in the south and central regions of Sweden. Control participants were randomly selected from a continuously updated national population register, with consideration given to age, sex, and living area. If the selected control was not traceable, reported having RA, or refused to participate, a new control was selected using the same procedure.

Sample set 4, obtained from the Genomics Collaborative, (GCI), comprised 475 RF-positive RA patients and 475 individually matched controls from the US and has been described in detail elsewhere [5]. In brief, all case samples were from white North Americans who were RF-positive and whose condition met the 1987 American College of Rheumatology diagnostic criteria for RA. Control samples were taken from a pool of healthy white individuals with no medical history of RA. A single control was matched to each case on the basis of sex, age ( $\pm 5$  y), and ethnicity (grandparental country/region of origin).

Patients from sample sets 1 and 4 had considerably longer disease duration at inclusion ( $13.8 \pm 10.1$  y and  $11.7 \pm 10.0$  y, respectively) as compared to patients from sample sets 2 and 3 ( $<3$  y) (Table S1). All controls were healthy unrelated white individuals originating from the same geographical area as the patients. There was no overlap between cases and controls across all studies. All protocols and recruitments were approved by national and/or local institutional review boards, and informed written consent was obtained from all participants.

### **SNP Selection and Genotyping**

We chose 40 polymorphisms spanning *TRAF1/C5* and their flanking genes *PHD finger protein 19 (PHF19)* and *centrosomal protein 110 kDa (CEP110)* for this study (Table 1). We selected haplotype tagging SNPs (htSNPs) from the International HapMap Project database (<http://www.hapmap.org/index.html>) as well as random SNPs from the University of California Santa Cruz database (<http://genome.ucsc.edu/>) to ascertain maximum haplotype information for each of the genes and intergenic regions that are likely to harbor regulatory regions. Chromosomal locations of the SNPs were extracted from SNPPER (<http://snpper.chip.org/>) Goldenpath hg17, dbSNP build 123. Genotyping across all studies was performed as described in detail in Protocol S1. Three of the 40 SNPs were excluded from further analysis after reviewing results from sample set 1, two of which (SNP13/rs4837803 and SNP27/rs10119768) were not polymorphic and one (SNP35/rs7856420) of which deviated from Hardy-Weinberg equilibrium.

### **Statistical Analysis**

#### **Single SNP and haplotype analysis**

Single SNP analysis and genetic model assessment was initially performed using SPSS version 12.0 (SPSS, <http://www.spss.com/>) in sample set 1. We did not find evidence of a recessive model for any SNPs. Since all SNPs were in Hardy-Weinberg equilibrium and adhered to the additive model of association, we performed further tests using allelic comparisons. Single- and multilocus allelic analyses were performed using Haploview version 3.32 (MIT, <http://www.broad.mit.edu/mpg/haploview/>) [18] with 40 SNPs in sample set 1 followed by six significant SNPs in sample set 2, the three most significant SNPs in sample set 3, and the single best-associating SNP in sample set 4. Odds ratios were calculated using Epi Info v6 (CDC, <http://www.cdc.gov/epiinfo/>). All *p*-values reported were two-sided. A *p*-value < 0.05 was considered significant. Based on the linkage disequilibrium (LD) structure in sample set 1, haplotype blocks were inferred under the algorithm of Gabriel et al. [19] in Haploview 3.32. To further minimize haplotypic uncertainty, we used the software TagSNPs version 1.0 (<http://www-rcf.usc.edu/~stram/tagSNPs.html>) to identify eight htSNPs from block 2 and ten htSNPs from block 3 (global  $R^2_h > 98\%$ ) [20]. The  $R^2_h$  coefficient is the squared correlation between the true haplotype count (number of copies of a haplotype) and the haplotype count predicted by TagSNPs. We chose htSNPs so that haplotypes were predicted with a global  $R^2_h$  value of 0.95 or above, indicating a high accuracy. Haplotype analyses using these htSNPs were performed in Haploview 3.32. Global *p*-values for haplotype associations were calculated using the software Haplo.Stats version 1.2 (<http://mayoresearch.mayo.edu/mayo>) used for estimating haplotype effects under the generalized linear model [21]. htSNPs from blocks 2 and 3 in sample set 1 were further investigated by sliding-window analysis in Haploview 3.32 to determine the basis of the associated haplotypes.

### **Combining datasets**

Odds ratios (OR)s from all sample sets were combined by the random model of the Cochran-Mantel-Haenszel test as implemented in EasyMA [22]. A Breslow-Day test of between-stage heterogeneity was also performed in EasyMA to test for consistency across sample sets [23]. We observed evidence of heterogeneity for SNP rs4836834 at  $p < 0.05$  (Table S2).

### **Logistics regression**

Forward conditional logistics regression was performed using all six significant SNPs from sets 1 and 2 in SPSS 12.0. Genotypes were coded as categorical variables 0, 1, and 2 with the nonassociated genotype as reference. No evidence of a recessive effect was observed, as detailed in Table S3.

### **Population attributable risk**

Population-attributable risk (PAR) was calculated using the population prevalence of the exposure (P) and the relative risk associated with the exposure (R), as follows [24].

$$\text{PAR} = \frac{P(R - 1)}{1 + P(R - 1)}$$

### **Autoantibodies**

Baseline laboratory parameters included IgM-RF (ELISA) and IgG-ACPA (ELISA, Immunoscan RA Mark2, Euro-Diagnostica, <http://www.eurodiagnostica.com/>). The cut-off for antibody positivity was set according to the manufacturer's instructions. Since baseline ACPA status was not available for most samples from the BEST study, we restricted our analysis to the EAC and EIRA cohorts. Autoantibody (RF and ACPA) and genotype status were available from 419 patients from the EAC (we additionally genotyped RF-negative and ACPA-negative patients in this cohort) and 1,395 patients from the EIRA study.

### **Severity**

Radiographs of hands and feet were scored at baseline and 2 y ( $n = 278$ ) using the Sharp-van der Heijde method (erosions and joint space narrowing of hands and feet) [25]. Data was available from 193 A carriers (AA+AG) and 85 non-A carriers (GG) from the EAC cohort. The prevalence of either RF (A carriers 0.9%, non-A carriers 56.5%,  $p = 0.485$ ) or anti-citrullinated protein antibodies (ACPA) (A carriers 60.7%, non-A carriers 60.8%,  $p = 0.986$ ) is not significantly different between the two groups analyzed. Differences in means between groups were calculated using sharp scores adjusted for baseline with a two-sided nonparametric Mann-Whitney test.

## Transcription Factor Binding Sites

Transcription factor binding sites (TFBSs) were predicted using MAPPER (<http://bio.chip.org/mapper/mapper-main>), a platform that combines information from two well-known TFBS databases, TRANSFAC and JASPAR. The prediction is generated from a hidden Markov model and is based on experimentally determined binding sites in multiple genomes [26].

## Results

### *The TRAF1/C5 Region Associates with RA*

To investigate whether the *TRAF1/C5* region on Chromosome 9q33-34 associates with RA, we selected a total of 40 polymorphisms spanning these candidates and their flanking genes for genotyping. Tagging SNPs as well as random SNPs were included to ascertain maximum haplotype information for each of the genes and to ensure coverage of intergenic regions which may harbor regulatory polymorphisms.

Single SNP analysis performed in the first set of RA patients ( $n = 290$ ) and controls ( $n = 254$ ) revealed significant association between six SNPs in the *TRAF1/C5* region (SNPs 4, 7, 10, 14, 15, and 16) and RA ( $p = 0.0104, 0.0153, 0.0080, 0.0039, 0.0039, \text{ and } 0.0250$ , respectively) (Table 1). One SNP in *PHF19* (SNP3/rs10985070) and one in *CEP110* (SNP36/rs10818503) also showed moderate association with RA ( $p = 0.0387$  and  $0.0257$ , respectively). None of the SNPs investigated showed evidence of a recessive mode of association.

To delineate the haplotypic architecture, we estimated the underlying haplotype block structure of the 32 SNPs with minor allele frequency  $> 5\%$  in the controls. We identified a potential recombination spot at SNP3/rs10985070 and SNP24/rs7026551 that divides the region into three inheritance blocks, blocks 1, 2, and 3. Block 1 is 8 kb and encompasses the 3' end to intron 9 of *PHF19*; block 2 extends from *TRAF1* through 24 kb of intergenic sequence to exon 32 of *C5*; while block 3 (178 kb) contains the remainder of the *C5* gene and *CEP110* (Figure 1A). The LD structure in the Dutch population was similar to the structure reported by the International HapMap Project (unpublished data).

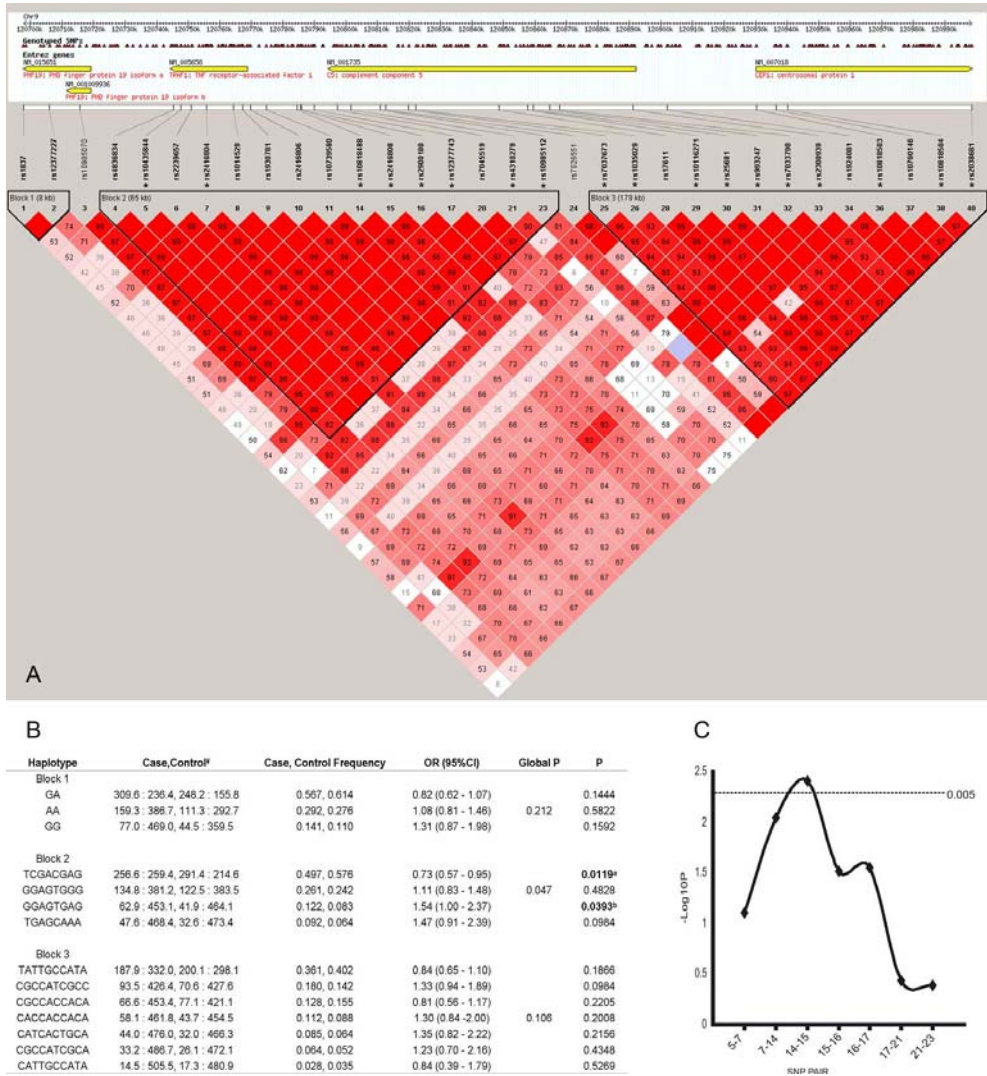


**Table 1.** Association of Candidate Gene SNPs with RA

SNP ID	SNP rs ID	Gene	Location	Chromosome Location <sup>a</sup>	Allele1/ Allele2	Susceptibility Allele	Allele Ratios <sup>b</sup> : Cases, Controls	Frequency <sup>c</sup> : Cases, Controls	Allelic OR (95% CI)	p-Value
1	rs1837	<i>PHF19</i>	3'UTR/intergenic	120697783	G/A	A	159:385, 110:290	0.292, 0.275	1.09 (0.81–1.47)	0.5611
2	rs12377227	<i>PHF19</i>	Intron 9	120706087	A/G	G	74:452, 44:354	0.141, 0.111	1.32 (0.87–2.00)	0.1742
3	rs10985070	<i>PHF19</i>	Intron 4	120715675	A/C	C	263:281, 168:236	0.483, 0.416	1.31 (1.01–1.72)	<b>0.0387</b>
4	rs4836834	<i>TRAF1</i>	Exon 8/3'UTR	120745455	A/T	T	260:268, 209:297	0.492, 0.413	1.38 (1.07–1.78)	<b>0.0104</b>
5	rs10435844	<i>TRAF1</i>	Intron 7	120747753	T/G	G	182:304, 158:326	0.374, 0.326	1.24 (0.94–1.62)	0.1169
6	rs2239657	<i>TRAF1</i>	Intron 7	120751074	A/G	G	207:329, 164:336	0.386, 0.328	1.29 (0.99–1.68)	0.0509
7	rs2416804	<i>TRAF1</i>	Intron 4	120755950	C/G	G	224:228, 206:288	0.496, 0.417	1.37 (1.05–1.79)	<b>0.0153</b>
8	rs1014529	<i>TRAF1</i>	Intron 3	120764497	G/C	C	208:332, 165:333	0.385, 0.331	1.26 (0.97–1.64)	0.0708
9	rs1930781	<i>TRAF1</i>	Intron 2	120767388	A/G	G	208:332, 166:336	0.385, 0.331	1.27 (0.98–1.65)	0.0668
10	rs2416806	<i>TRAF1</i>	5'UTR/intergenic	120769846	C/G	G	214:308, 160:326	0.410, 0.329	1.42 (1.09–1.85)	<b>0.0080</b>
11	rs10739580	<i>TRAF1</i>	5'UTR/intergenic	120774836	T/C	C	192:314, 154:320	0.379, 0.325	1.27 (0.97–1.67)	0.0741
12	rs6478491	<i>TRAF1</i>	5'UTR/intergenic	120778044	G/C	C	8:532, 7:493	0.015, 0.014	1.06 (0.35–3.26)	0.9123
14	rs10818488	<i>TRAF1</i>	5'UTR/intergenic	120784641	G/A	A	247:255, 193:289	0.492, 0.400	1.45 (1.12–1.88)	<b>0.0039</b>
15	rs2416808	<i>TRAF1</i>	5'UTR/intergenic	120785837	A/G	G	247:255, 193:289	0.492, 0.400	1.45 (1.12–1.88)	<b>0.0039</b>
16	rs2900180	<i>TRAF1</i>	5'UTR/intergenic	120785936	C/T	T	216:332, 161:331	0.394, 0.327	1.34 (1.03–1.74)	<b>0.0250</b>
17	rs12377743	<i>C5</i>	3'UTR/intergenic	120790004	G/A	A	58:482, 39:461	0.107, 0.078	1.42 (0.91–2.22)	0.1032
18	rs7874632	<i>C5</i>	3'UTR/intergenic	120792886	G/C	C	4:512, 2:500	0.008, 0.004	1.95 (0.31–15.38)	0.4323
19	rs12005198	<i>C5</i>	3'UTR/intergenic	120793949	A/G	A	539:1, 501:1	0.998, 0.998	1.08 (0.00–39.38)	0.9588
20	rs7045519	<i>C5</i>	Intron 39	120797394	C/A	A	58:482, 40:462	0.107, 0.080	1.39 (0.89–2.17)	0.1255
21	rs4310279	<i>C5</i>	Intron 38	120801702	A/G	G	148:392, 124:374	0.274, 0.249	1.11 (0.84–1.48)	0.3587
22	rs16910233	<i>C5</i>	Intron 37	120803165	C/G	G	4:518, 3:499	0.008, 0.006	1.28 (0.24–7.24)	0.7433
23	rs10985112	<i>C5</i>	Intron 32	120810962	G/A	A	54:486, 33:465	0.100, 0.066	1.57 (0.98–2.52)	0.0501
24	rs7026551	<i>C5</i>	Intron 31	120812687	A/C	C	121:393, 94:402	0.235, 0.190	1.32 (0.96–1.80)	0.0749
25	rs7037673	<i>C5</i>	Intron 28	120820038	C/T	C	319:195, 279:219	0.621, 0.560	1.28 (0.99–1.66)	0.0508
26	rs1035029	<i>C5</i>	Intron 27	120822372	A/G	G	206:316, 188:316	0.395, 0.373	1.10 (0.85–1.42)	0.4766
28	rs17611	<i>C5</i>	Exon 19	120848754	G/A	G	251:187, 267:227	0.583, 0.540	1.19 (0.91–1.55)	0.1935
29	rs10116271	<i>C5</i>	Intron 15	120857702	C/T	C	255:251, 221:263	0.504, 0.457	1.21 (0.93–1.56)	0.1362
30	rs25681	<i>C5</i>	Exon 13	120859559	C/T	C	297:205, 265:227	0.592, 0.539	1.24 (0.96–1.61)	0.0918
31	rs993247	<i>C5</i>	Intron 10	120864803	A/G	A	307:213, 274:226	0.590, 0.548	1.19 (0.92–1.54)	0.1717
32	rs7033790	<i>C5</i>	Intron 8	120867946	C/T	T	129:389, 100:402	0.249, 0.199	1.33 (0.98–1.81)	0.0565
33	rs2300939	<i>C5</i>	Intron 1	120891271	C/T	T	44:454, 34:452	0.088, 0.070	1.29 (0.79–2.11)	0.2856
34	rs1924081	<i>C5</i>	5'UTR/intergenic	120902001	T/A	A	137:413, 100:398	0.249, 0.201	1.32 (0.98–1.79)	0.0620
36	rs10818503	<i>CEP110</i>	Exon 1	120930324	A/G	G	181:333, 144:358	0.352, 0.287	1.35 (1.03–1.78)	<b>0.0257</b>
37	rs10760146	<i>CEP110</i>	Intron 2	120936639	C/T	C	332:216, 276:220	0.606, 0.556	1.23 (0.95–1.58)	0.1061
38	rs10818504	<i>CEP110</i>	Exon 5	120940243	C/T	C	312:210, 280:226	0.598, 0.553	1.20 (0.93–1.55)	0.1504
39	rs17292952	<i>CEP110</i>	Exon 16	120983394	G/A	G	525:25, 473:25	0.955, 0.950	1.11 (0.61–2.03)	0.7189
40	rs2038681	<i>CEP110</i>	Intron 25	120998594	A/C	C	94:416, 74:428	0.184, 0.147	1.31 (0.92–1.85)	0.1147

Data in bold indicate significant SNPs.

- a. Chromosomal locations were extracted from SNPPER Goldenpath hg17, dbSNP build 123.
- b. Number of alleles were compared in cases versus controls: allele1:allele2 cases, allele1:allele2 controls.
- c. Frequency of the susceptibility allele.



**Figure 1.** LD Structure and Haplotype Association across the TRAF1/C5 Region in Sample Set 1 (290 RA Patients and 254 Controls)

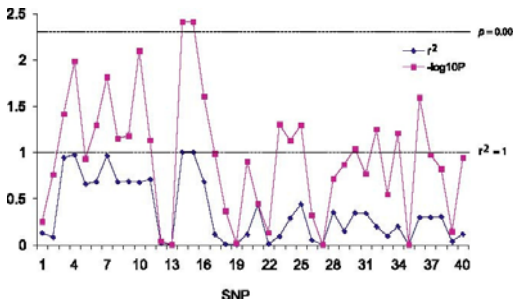
(A) Haplotype block structure was predicted on the basis of the strength of pairwise LD, which is presented as a 2x2 matrix; red represents very high LD. (D), white indicates absence of correlation between SNPs, and blue indicates high correlation with a low level of significance. SNPs that were chosen for haplotype analysis are indicated along the top by an asterisk.

(B) Using htSNPs from each block, indicated by the asterisk in (A), haplotypes were inferred with a certainty of above 98% as represented by the  $R^2_h$  value. <sup>a</sup>Comparisons were made between one haplotype versus all others in cases and controls. <sup>a</sup>The protective haplotype that is significantly less frequent in cases. <sup>b</sup>The susceptible haplotype that is significantly more frequent in patients.

(C) Sliding window of the susceptible haplotype using consecutive two-SNP combinations of the htSNPs reveals that SNP14 and SNP15 account for most of the association observed. For each of the SNP pairs we show the  $-\log_{10} p$ -value. The dotted line indicates a nominal  $p$ -value of 0.005.

To further minimize haplotypic uncertainty and to identify the minimal combination of SNPs that provide maximum information content within each block, we scanned these two blocks independently using the software TagSNPs. In block 2 we identified a minimal set of eight htSNPs with a global  $R^2_h$  of 0.996, and in block 3 we found ten htSNPs with an  $R^2_h$  of 0.985, indicating that haplotypes can be inferred with >95% certainty. Haplotypes were predicted, and analyses from all blocks revealed that the association with RA was restricted to SNPs in block 2 (Figure 1B), as demonstrated by the global  $p$ -value of association ( $p < 0.05$ ) and suggesting the possible involvement of *TRAF1* and/or the 3' end of *C5*.

Of the four common haplotypes capturing > 95% of the variation, two significantly associated haplotypes were observed, one increased in RA patients (susceptible haplotype  $p = 0.039$ ), and one over-represented in controls (protective haplotype  $p = 0.012$ ). By applying a two-marker sliding window analysis, we observed a significant peak centered on SNP14/rs10818488 and SNP15/rs2416808 ( $p = 0.0039$ ) (Figure 1C). Using three- and five-marker windows did not alter the outcome, suggesting that the significance seen with the other *TRAF1* SNPs (Table 1) may be due to LD. To explore this possibility, we analyzed the  $r^2$ -values with respect to SNP14. We confirmed that the most significant SNPs, which are located in *TRAF1* and the adjacent intergenic region, are highly correlated with SNP14 ( $r^2 > 0.90$ ) (Figure 2).



**Figure 2.** RA-Associated SNPs are Highly Linked to SNP14

Pairwise LD ( $r^2$ ) between associated SNP14/rs10818488 and all other SNPs genotyped was calculated. For each of the SNPs listed along the x-axis we show the  $-\log_{10}$  (y-axis) of the  $p$ -values for RA patients versus controls. Dotted lines indicate a nominal  $p$ -value of 0.005 and a maximal  $r^2$  value of 1. In a logistics regression model, only SNP14/rs10818488 remained statistically significant ( $p = 6.16 \times 10^{-4}$ ).

### **Replication in Three Independent Sample Sets from The Netherlands, Sweden, and the US**

Six tagging SNPs which were significant ( $p < 0.05$ ) in the initial study were selected for replication in a fully independent set of Dutch cases and controls (sample set 2). Of these, only SNP14/rs10818488 and SNP15/rs2416808 were statistically significant ( $p < 0.05$ ). Haplotype and a two-marker sliding window analysis localized the strongest region of association to SNP14 and SNP15, confirming the results from sample set 1 (unpublished data). Combined analyses of the data from sample sets 1 and 2 showed an even stronger association for SNP14 (OR 1.34, 95% confidence interval [CI] 1.13–1.58;  $p = 5.56 \times 10^{-4}$ ) and SNP15 (OR 1.33, 95% CI 1.13–1.57;  $p = 6.65 \times 10^{-4}$ ) (Table 2). Although the other four SNPs did not reach statistical significance in sample set 2, they were highly significant in the combined analysis (Table 2). To evaluate putative modes of inheritance, we calculated genotype-specific ORs in the combined dataset as detailed in Table S3. All SNPs were consistent with an additive model. On the basis

of forward conditional logistics regression, SNP14 remained in the model with a heterozygote (AG) OR of 1.38 (95% CI 1.04–1.83,  $p = 0.027$ ) and a homozygote (AA) OR of 2.06 (95% CI 1.42–2.98,  $p = 1.29 \times 10^{-3}$ ).

Similar replication of three SNPs in a cohort of Swedish patients and controls (sample set 3) confirmed association with SNP14/rs10818488 ( $p = 0.0078$ ) (Table 2). A combined analysis of patients and controls of European origin (Dutch and Swedish) with 2,244 RA patients and 1,524 controls (sample sets 1, 2, and 3) showed that the most significant associations could again be attributed to SNP14 (OR 1.24, 95% CI 1.11–1.38,  $p = 1.73 \times 10^{-5}$ ) and SNP15/rs2416808 (OR 1.23, 95% CI 1.09–1.40,  $p = 7.21 \times 10^{-5}$ ) (Table 2). Additionally, these findings were further replicated in a case-control sample set from the US. Since LD analysis in the original three sample sets showed SNP14 and SNP15 to be highly correlated ( $r^2 > 0.98$ ) we genotyped only SNP14 and confirmed that the minor susceptibility allele was associated with RA risk (OR 1.36, 95% CI 1.13–1.64;  $p = 0.001$ ) (Table 2). Combined analysis of SNP14 in all four independent sets ( $(n_{\text{cases/controls}} = 2,719/1,999)$ ) yielded a highly significant association  $OR_{\text{common}} = 1.26$ , 95% CI 1.15–1.37,  $p_{\text{combined}} = 1.40 \times 10^{-8}$  and a PAR of 6.1% (95% CI 4.0–8.5).

**Table 2.** Association of Significant SNPs in Four Independent Sample Sets

SNP	Set 1 (Netherlands)			Set 2 (Netherlands)			Sets 1 + 2		
	Allele Ratios <sup>a</sup> Cases, Controls	Allelic OR (95% CI)	p- Value	Allele Ratios <sup>a</sup> Cases, Controls	Allelic OR (95% CI)	p- Value	Allele Ratios <sup>a</sup> Cases, Controls	Allelic OR (95% CI) <sup>b</sup>	p- Value
SNP4/rs4836834	260:268, 209:297	1.38 (1.07–1.78)	0.0104	378:432, 225:301	1.17 (0.93–1.47)	0.1626	638:700, 434:598	1.41 (0.96–2.13)	0.078
SNP7/rs2416804	224:228, 206:288	1.37 (1.05–1.79)	0.0153	369:415, 224:300	1.19 (0.95–1.50)	0.1242	593:643, 430:588	1.27 (1.07–1.50)	0.006
SNP10/rs2416806	214:308, 160:326	1.42 (1.09–1.85)	0.0080	296:489, 180:340	1.14 (0.90–1.45)	0.2559	510:798, 340:666	1.26 (1.02–1.56)	0.029
SNP14/rs10818488	247:255, 193:289	1.45 (1.12–1.88)	0.0039	387:425, 220:304	1.26 (1.00–1.58)	0.0419	634:680, 413:593	1.34 (1.13–1.58)	<b>5.56 × 10<sup>-4</sup></b>
SNP15/rs2416808	247:255, 193:289	1.45 (1.12–1.88)	0.0039	387:425, 221:303	1.25 (0.99–1.57)	0.0494	634:680, 414:592	1.33 (1.13–1.58)	<b>6.65 × 10<sup>-4</sup></b>
SNP36/rs10818503	181:333, 144:358	1.35 (1.03–1.78)	0.0257	259:527, 166:352	1.04 (0.82–1.33)	0.7329	440:860, 310:710	1.17 (0.98–1.40)	0.210
SNP	Set 3 (Sweden)			Sets 1 + 2 + 3					
	Allele Ratios <sup>a</sup> Cases, Controls	Allelic OR (95% CI)	p- Value	Allele Ratios <sup>a</sup> Cases, Controls	Allelic OR (95% CI) <sup>b</sup>	p- Value			
SNP4/rs4836834	—	—	—	—	—	—			
SNP7/rs2416804	—	—	—	—	—	—			
SNP10/rs2416806	1,212:1,756, 677:1,089	1.11 (0.98–1.25)	0.0893	1,722:2,554, 1,017:1,755	1.18 (1.03–1.34)	0.015			
SNP14/rs10818488	1,500:1,500, 840:984	1.17 (1.04–1.32)	<b>0.00781</b>	2,134:2,180, 1,253:1,577	1.23 (1.12–1.36)	<b>1.73 × 10<sup>-5</sup></b>			
SNP15/rs2416808	1,351:1,353, 772:890	1.15 (1.02–1.30)	<b>0.0241</b>	1,985:2,033, 1,186:1,482	1.23 (1.09–1.40)	<b>7.21 × 10<sup>-5</sup></b>			
SNP36/rs10818503	—	—	—	—	—	—			
SNP	Set 4 (US)			Sets 1 + 2 + 3 + 4					
	Allele Ratios <sup>a</sup> Cases, Controls	Allelic OR (95% CI)	p- Value	Allele Ratios <sup>a</sup> Cases, Controls	Allelic OR (95% CI) <sup>b</sup>	p- Value			
SNP4/rs4836834	—	—	—	—	—	—			
SNP7/rs2416804	—	—	—	—	—	—			
SNP10/rs2416806	—	—	—	—	—	—			
SNP14/rs10818488	427:513, 356:582	1.36 (1.13–1.64)	<b>0.00102</b>	2,561:2,693, 1,609:2,159	1.26 (1.15–1.37)	<b>1.40 × 10<sup>-8</sup></b>			
SNP15/rs2416808	—	—	—	—	—	—			
SNP36/rs10818503	—	—	—	—	—	—			

Data in bold indicate the p-value of the most significant SNPs.

a. The allele frequencies between allele1:allele2 in cases was compared to allele1:allele2 in controls. Allele 1 refers to the susceptibility allele from Table 1.

b. Mantel-Haenzel OR as calculated under the random model. Raw data can be obtained from Table S2.

### Association with Autoantibody-Positive Disease

RA is a heterogeneous disease with a considerable variation in phenotype as evidenced by the fact that some patients are autoantibody-positive whereas others are not. Antibodies to citrullinated protein antigens, called ACPAs, have gained much interest as current data suggest

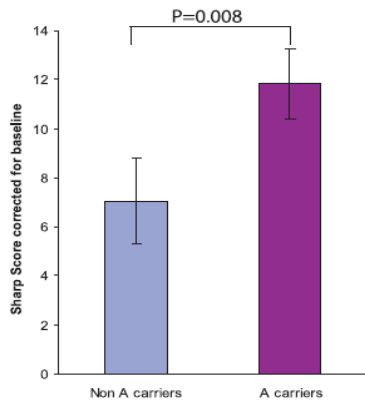
that ACPA-positive and negative RA may have different genetic risk factors [27]. To investigate whether the *TRAF1/C5* region is associated with a specific phenotype of RA, we next stratified patients for autoantibody status from whom baseline ACPA and RF measurements were available ( $n = 1,814$ ). Interestingly, SNP14/rs10818488 mainly predisposes to autoantibody-positive disease when compared to controls (OR 1.25, 95% CI 1.11–1.40,  $p = 2.27 \times 10^{-4}$ ) (Figure 3A). Although we also observed an increase in the frequency of the A allele in autoantibody-positive as compared to autoantibody-negative disease, this difference did not reach formal statistical significance (OR 1.15, 95% CI 0.98–1.34,  $p = 0.0789$ ). These data therefore suggest that the current genetic risk factor may be predominant in the autoantibody-positive subset of RA patients.

### **Association with Severity**

Because the clinical course of RA can vary considerably ranging from nonerosive disease to rapidly progressive joint damage, we also analyzed whether the SNPs in the *TRAF1/C5* region were involved with RA progression. Annual X-rays of the hands and feet of patients were assigned Sharp–van der Heijde units, a combined score for bone erosions and joint space narrowing. Carriers of the minor susceptibility allele of SNP14/rs10818488 had an almost 2-fold higher severe disease course at 2 y after inclusion as compared to the non-A carriers (Figure 3B; mean  $\pm$  SE score of A carriers/non-A carriers,  $11.4 \pm 1.4/7.1 \pm 1.8$ ;  $p = 0.008$ ) indicating that the A allele predisposes not only to RA susceptibility, but also to severity.

	Number of Alleles			Allele Frequency		Cases vs Controls		AutoAntibody Positive vs Negative	
	A	G	Total	A	G	Allelic OR (95% CI)	P	Allelic OR (95% CI)	P
Controls*	1253	1577	2830	0,44	0,56	-	-	-	-
ACPA-RF-	482	558	1040	0,46	0,54	1,09 (0,94 - 1,26)	0,2509	-	-
ACPA+RF-	120	136	256	0,47	0,53	1,11 (0,85 - 1,45)	0,4228	1,02 (0,77 - 1,36)	0,8792
RF+ACPA-	216	224	440	0,49	0,51	1,21 (0,99 - 1,49)	0,0588	1,12 (0,89 - 1,40)	0,3336
RF+ACPA+	941	951	1892	0,50	0,50	1,25 (1,11 - 1,40)	<b>2,27x10<sup>-6</sup></b>	1,15 (0,98 - 1,34)	<b>0,0789</b>

\* Controls refer to the combined controls from sets 1, 2 and 3.



**Figure 3.** The A Allele of SNP14 Is Associated with Distinct Phenotypes of RA. (Top, table) The frequency of the A allele of patients from whom baseline autoantibody status (ACPA and RF) were available was calculated. ORs and pvalues were calculated between each subgroup and controls, and indicated a predominantly higher frequency of the A allele in ACPA- and RF-positive patients. (Bottom, bar graph) Progression of joint damage in Sharp–van der Heijde units (“Sharp score”) is higher in the presence of the minor A allele of SNP14. Radiological data of 193 A carriers and 85 non-A carriers were available, and differences between the groups were calculated based on disease severity after 2 y corrected for baseline.

### **SNP14/rs 10818488 Is Located in a TFBS**

To investigate the potential functional effect of this SNP, we scanned for transcription factor binding sites using MAPPER [26]. The SNP14 susceptibility A allele encodes a potential binding site for EP300, a histone acetyl transferase that regulates transcription via chromatin remodeling. In the absence of this allele, the binding of EP300 to this region is predicted to be disrupted, potentially disturbing the epigenetic tag for transcriptional activation. We hypothesize that this putative transcription factor binding site may be involved in the regulation of the neighbouring *TRAF1* and/or *C5* gene.

## Discussion

Using a candidate-gene approach, we identified the *TRAF1/C5* region on Chromosome 9q33-34 as a susceptibility and severity factor for RA. This region was also associated with RA in a large-scale genetic association study (Schrodi et al., unpublished data). It is, therefore, intriguing to see that these independent studies, in which the process leading to results differed, give similar results, and in doing so provide strong evidence for the *TRAF1/C5* region as a true RA-associated genetic variant. The recent genome-wide study performed by the Wellcome Trust failed to identify this region as a candidate for RA [28]. Although it is difficult to speculate why this region was not detected in the Wellcome Trust Case Control Consortium study, we do note that none of the SNPs showing strong association in our hands was genotyped by the Wellcome Trust. Additionally, in line with our finding that this genetic risk factor is predominant in the autoantibody-positive subgroup, substratifications of the specific RA phenotypes may be needed to detect significant association.

The protein encoded by *TRAF1* is a member of the TNF receptor-associated factor (TRAF) protein family, which associates with and mediates the signal transduction from various receptors of the TNF receptor superfamily, including the receptor for TNF $\alpha$  [29]. In addition to a direct role in TNF $\alpha$  signaling, *TRAF1* has also been implicated in the activation and proliferation of T cells [30] and is expressed ubiquitously by other cells of the immune system including monocytes and B cells [31]. It is therefore possible that TRAF1 could play a role in RA by aiding the maintenance of the proinflammatory environment. Likewise, studies have also shown that perpetuation of inflammation coincides with increased levels of the anaphylatoxin C5a in the synovial fluid of RA patients [9]. Further studies in mice identifying *C5* as a candidate gene and showing that *C5* deficiency results in lower incidence and less-severe disease course support the role of this gene in inflammation [7,11]. It is therefore likely that although the primary function of the complement system is to protect the host from microorganisms, a deregulated activity of its central component, *C5*, can play a substantial role in inflammatory diseases as well.

In order to capture the variation within these candidate genes and potential regulatory regions, we genotyped both SNPs that were intragenic and those located 5' and 3' of the genes. Interestingly, the strongest replicated association was observed with SNP14/rs10818488, which maps to an intergenic region ~10 kb from both *TRAF1* and *C5* and is present in a TFBS which may regulate the transcriptional activity of its neighbouring genes. However, formal testing of all known variation within this region, both genetic and biological, will be necessary to pinpoint the precise biological process that is altered by the RA-associated variant(s) present in this region.

We found a strong association of this region in all four independent sample sets which represent varying disease durations (<3 to >10 y) as well as a correlation with disease progression. More importantly, these phenotypic data on joint destruction not only indicate that the *TRAF1/C5* region predisposes to RA, but also suggest that within the RA population, patients harboring the minor susceptibility allele of SNP14/rs10818488 tend to experience a more severe disease course. Although the above findings most likely exclude the possibility of a spurious association, especially since each case group was assigned an ethnically and geographically matched control group, background levels of population stratification as described by Cardon et al. [32]

may exist in the different populations under study. It is therefore conceivable that the slight variation in the observed effect between the four populations may partially be due to varying sample sizes and background levels of population admixture. RA is a common complex disease that results from the interaction of multiple genetic variants, each with relatively low penetrance, with an array of environmental triggers [33]. In advance of a genetic profile that can accurately pinpoint individuals at risk, identification of these genetic variants can provide insight into the underlying mechanisms of disease and the specific pathways associated with disease induction and/or progression. Understanding the function of these common disease-associated variants will be important to identify potential targets for intervention strategies that could prove useful to all patients, whether or not they carry the disease-associated variant.

In summary, this study provides robust evidence from four independent sample sets (two of Dutch origin, one of Swedish descent, and one from the US) that genetic variants within the *TRAF1/C5* region are associated with RA, indicating a possible role for these immune-related genes in the biological process underlying RA disease pathogenesis.

### **Limitations of This Study**

Our study defines the *TRAF1/C5* as a novel genetic region present in the human genome that predisposes to RA. However, the causative variation (SNP) or the biological mechanism explaining this association is not yet known. Although it could be that the current identified polymorphism is causative, other proxies in high LD with this SNP could also be responsible for this issue, which can be addressed in more detailed by functional studies. Furthermore, although our data indicate a predominant association with autoantibody-positive disease, our study is underpowered to exclude the possibility that the *TRAF1/C5* region also predisposes to autoantibody-negative disease. By combining information obtained from other cohorts in which both the autoantibody and the *TRAF1/C5* status are known, this question should be resolved in the future. As it has been indicated that distinct genetic risk factors underlie either autoantibody-positive or autoantibody-negative disease, such additional information would provide more detailed knowledge on the genetic heterogeneity of these two distinct phenotypes of RA.



## References

- (1) Firestein GS (2003) Evolving concepts of rheumatoid arthritis. *Nature* 423: 356–361.
- (2) Seldin MF, Amos CI, Ward R, Gregersen PK (1999) The genetics revolution and the assault on rheumatoid arthritis. *Arthritis Rheum* 42: 1071–1079.
- (3) MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, et al. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 43: 30–37.
- (4) Deighton CM, Walker DJ, Griffiths ID, Roberts DF (1989) The contribution of HLA to rheumatoid arthritis. *Clin Genet* 36: 178–182.
- (5) Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, et al. (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 75: 330–337.
- (6) Maini RN (2004) Current and new antitumor necrosis factor agents in perspective. *Arthritis Res Ther* 6: S1–S2.
- (7) Wang Y, Kristan J, Hao L, Lenkoski CS, Shen Y, et al. (2000) A role for complement in antibody-mediated inflammation: C5-deficient DBA/1 mice are resistant to collagen-induced arthritis. *J Immunol* 164: 4340–4347.
- (8) Wang Y, Rollins SA, Madri JA, Matis LA (1995) Anti-C5 monoclonal antibody therapy prevents collagen-induced arthritis and ameliorates established disease. *Proc Natl Acad Sci U S A* 92: 8955–8959.
- (9) Hogasen K, Mollnes TE, Harboe M, Gotze O, Hammer HB, et al. (1995) Terminal complement pathway activation and low lysis inhibitors in rheumatoid arthritis synovial fluid. *J Rheumatol* 22: 24–28.
- (10) Grant EP, Picarella D, Burwell T, Delaney T, Croci A, et al. (2002) Essential role for the C5a receptor in regulating the effector phase of synovial infiltration and joint destruction in experimental arthritis. *J Exp Med* 196: 1461–1471.
- (11) Ji H, Gauguier D, Ohmura K, Gonzalez A, Duchatelle V, et al. (2001) Genetic influences on the end-stage effector phase of arthritis. *J Exp Med* 194: 321–330.
- (12) Eyre S, Barton A, Shephard N, Hinks A, Brintnell W, et al. (2004) Investigation of susceptibility loci identified in the UK rheumatoid arthritis whole-genome scan in a further series of 217 UK affected sibling pairs. *Arthritis Rheum* 50: 729–735.
- (13) Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, et al. (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 31: 315–324.
- (14) Brinkman BM, Huizinga TW, Kurban SS, van der Velde EA, Schreuder GM, et al. (1997) Tumour necrosis factor alpha gene polymorphisms in rheumatoid arthritis: association with susceptibility to, or severity of, disease? *Br J Rheumatol* 36: 516–521.
- (15) Goekoop-Ruiterman YP, Vries-Bouwstra JK, Allaart CF, van Zeben D, Kerstens PJ, et al. (2005) Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): A randomized, controlled trial. *Arthritis Rheum* 52: 3381–3390.
- (16) van Aken J, van Bilsen JH, Allaart CF, Huizinga TW, Breedveld FC (2003) The Leiden Early Arthritis Clinic. *Clin Exp Rheumatol* 21: S100–S105.

- (17) Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L (2004) 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.
- (18) Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265.
- (19) Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. (2002) The structure of haplotype blocks in the human genome. *Science* 296: 2225–2229.
- (20) Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, et al. (2003) Choosing haplotype-tagging SNPs based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. *Hum Hered* 55: 27–36.
- (21) Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70: 425–434.
- (22) Cucherat M, Boissel JP, Leizorovicz A, Haugh MC (1997) EasyMA: A program for the meta-analysis of clinical trials. *Comput Methods Programs Biomed* 53: 187–190.
- (23) Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Metaanalysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33: 177–182.
- (24) Breslow NE, Day NE (1980) Statistical methods in cancer research. Volume I—The analysis of case-control studies. *IARC Sci Publ* 5–338.
- (25) van der Heijde DM (1996) Plain X-rays in rheumatoid arthritis: Overview of scoring methods, their reliability and applicability. *Baillieres Clin Rheumatol* 10: 435–453.
- (26) Marinescu VD, Kohane IS, Riva A (2005) MAPPER: A search engine for the computational identification of putative transcription factor binding sites in multiple genomes. *BMC Bioinformatics* 6: 79.
- (27) van Gaalen F, Ioan-Facsinay A, Huizinga TW, Toes RE (2005) The devil in the details: The emerging role of anticitrulline autoimmunity in rheumatoid arthritis. *J Immunol* 175: 5575–5580.
- (28) The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661–678.
- (29) Wajant H, Henkler F, Scheurich P (2001) The TNF-receptor-associated factor family: Scaffold molecules for cytokine receptors, kinases and their regulators. *Cell Signal* 13: 389–400.
- (30) Sabbagh L, Srokowski CC, Pule G, Snell LM, Sedgmen BJ, et al. (2006) A critical role for TNF receptor-associated factor 1 and Bim down-regulation in CD8 memory T cell survival. *Proc Natl Acad Sci U S A* 103: 18703–18708.
- (31) Zapata JM, Krajewska M, Krajewski S, Kitada S, Welsh K, et al. (2000) TNFR-associated factor family protein expression in normal tissues and lymphoid malignancies. *J Immunol* 165: 5084–5096.
- (32) Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. *Lancet* 361: 598–604.
- (33) Kallberg H, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, et al. (2007) 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.

## Supporting Information

### **Protocol S1.** Genotyping Methods

Found at doi:10.1371/journal.pmed.0040278.sd001 (26 KB DOC).

### **Table S1.** Clinical Characteristics of RA Patients from Four Independent Sample Sets

Found at doi:10.1371/journal.pmed.0040278.st001 (21 KB XLS).

### **Table S2.** Mantel-Haenzel OR and P for Combined Sets

Found at doi:10.1371/journal.pmed.0040278.st002 (22 KB XLS).

### **Table S3.** Genotypic Odds Ratios for the Most Significant SNPs in Sample Sets 1 and 2

Found at doi:10.1371/journal.pmed.0040278.st003 (21 KB XLS).

### **Table S4.** Primers and Conditions for PCR\_RFLP

Found at doi:10.1371/journal.pmed.0040278.st004 (21 KB XLS).

## Accession Numbers

The GenBank (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>) accession numbers for the genes discussed in this article are C5 (NM\_001735); CEP110 (NM\_007018); PHF19 (XM\_045308); TRAF1 (NM\_005658). The ExPASy UniProtKB/Swiss-Prot (<http://www.expasy.org/sprot/>) accession number for EP300 is Q09472.