Associations between Human leukocyte antigen, PTPN22, CTLA4 genotypes and rheumatoid arthritis phenotypes of autoantibody status, age at diagnosis and erosions in a large cohort study

E W Karlson, L B Chibnik, J Cui, R M Plenge, R J Glass, N E Maher, A Parker, R Roubenoff, E Izmailova, J S Coblyn, M E Weinblatt and N A Shadick

Ann Rheum Dis 2008;67;358-363; originally published online 31 Jul 2007; doi:10.1136/ard.2007.071662

Associations between Human leukocyte antigen, PTPN22, CTLA4 genotypes and rheumatoid arthritis phenotypes of autoantibody status, age at diagnosis and erosions in a large cohort study

E W Karlson,[1] L B Chibnik,[1] J Cui,[1] R M Plenge,[1] R J Glass,[1] N E Maher,[1] A Parker,[2] R Roubenoff,[2] E Izmailova,[2] J S Coblyn,[1] M E Weinblatt,[1] N A Shadick[1]

[1] Division of Rheumatology, Allergy, and Immunology, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; [2] Millenium Pharmaceuticals, Cambridge, Massachusetts, USA

Correspondence to: E W Karlson, 75 Francis Street, Boston, MA 02030, USA; ekarlson@partners.org

Accepted 14 July 2007
Published Online First 31 July 2007

ABSTRACT

Background: *HLA-DRB1* shared epitope (*HLA-SE*), *PTPN22* and *CTLA4* alleles are associated with cyclic citrullinated peptide (CCP) and rheumatoid arthritis (RA).

Objective: We examined associations between *HLA-SE, PTPN22*, *CTLA4* genotypes and RA phenotypes in a large cohort to (a) replicate prior associations with CCP status, and (b) determine associations with radiographic erosions and age of diagnosis.

Methods: A total of 689 RA patients from the Brigham RA Sequential Study (BRASS) were genotyped for *HLA-SE, PTPN22* (rs2476601) and *CTLA4* (rs3087243). Association between genotypes and CCP, rheumatoid factor (RF) erosive phenotypes and age at diagnosis were assessed with multivariable models adjusting for age, sex and disease duration. Novel causal pathway analysis was used to test the hypothesis that genetic risk factors and CCP are in the causal pathway for predicting erosions.

Results: In multivariable analysis, presence of any *HLA-SE* was strongly associated with CCP+ (odds ratio (OR) 3.05, 95% CI 2.18–4.25), and RF+ (OR 2.53, 95% CI 1.83–3.5) phenotypes; presence of any *PTPN22* T allele was associated with CCP+ (OR 1.81, 95% CI 1.24–2.66) and RF+ phenotypes (OR 1.84, 95% CI 1.27–2.66). CTLA4 was not associated with CCP or RF phenotypes. While *HLA-SE* was associated with erosive RA phenotype (OR 1.52, 95% CI 1.01–2.17), this was no longer significant after conditioning on CCP. *PTPN22* and *CTLA4* were not associated with erosive phenotype. Presence of any *HLA-SE* was associated with an average 3.6 years earlier diagnosis compared with absence of *HLA-SE* (41.3 vs 44.9 years, $p = 0.002$) and *PTPN22* was associated with a 4.2 years earlier age of diagnosis (39.5 vs 43.6 years, $p = 0.002$). *CTLA4* genotypes were not associated with age at diagnosis of RA.

Conclusions: In this large clinical cohort, we replicated the association between *HLA-SE* and *PTPN22*, but not *CTLA4* with CCP+ and RF+ phenotypes. We also found evidence for associations between HLA-SE, and PTPN22 and earlier age at diagnosis. Since *HLA-SE* is associated with erosive phenotype in unconditional analysis, but is not significant after conditioning on CCP, this suggests that CCP is in the causal pathway for predicting erosive phenotype.

Genetic factors are thought to be responsible for up to 50%–60% of rheumatoid arthritis (RA) risk.[1–3] Two genes have been unequivocally associated with RA susceptibility (Human leukocyte antigen *HLA-DRB1* and *PTPN22*), while other genes demonstrate strong, but inconclusive risk (eg. *CTLA*, *PADI4*).[4] Although the HLA associations with RA are complex,[5 6] the majority of the genetic signal from HLA is explained by alleles at the *HLA-DRB1* locus,[7] and account for approximately 30% of the genetic risk of RA.[1] In individuals of European ancestry, the associated *HLA-DRB1* alleles share a region of sequence similarity or "shared epitope" at amino acid positions 70–74 in the third hypervariable region of the *HLA-DRB1* molecule[1] (Human leukocyte antigen shared epitope; *HLA-SE*). Outside HLA, the only genetic polymorphism that has been associated with RA susceptibility in populations of European ancestry and replicated across multiple independent studies is *PTPN22*.[3 8–14] A missense allele (C→T) is associated with an increased risk of RA (rs2476601), with a summary odds ratio (OR) of 1.68 (1.53–1.84) from a meta-analysis.[15] The *HLA-SE* alleles and the *PTPN22* allele are more strongly associated with the phenotype cyclic citrullinated peptide positive (CCP+) RA.[4 16–19] Cytotoxic T lymphocyte-associated antigen 4 (*CTLA4*) gene has an A→G single nucleotide polymorphism (SNP) in the 3′ untranslated region (rs3087243), that is associated with increased risk of RA in several populations,[20–22] although the OR is much more modest (OR 1.20). In a large replication study *CTLA4* was more strongly associated with CCP+ RA.[4]

Phenotyping of disease subgroups in rheumatoid arthritis (RA), such CCP status, is important in studies of RA genetic and environmental epidemiology.[17 23–25] Environmental risk factors for RA differ in CCP+ and CCP– subgroups.[25] *HLA-SE* demonstrates a significant gene-environment interaction with cigarette smoking for susceptibility to CCP+ RA[23 26] and for CCP antibody status in RA[24] but not for CCP– RA.

Studies of genetic predictors of individual RA phenotypes have suggested associations between several genes and autoantibody status (CCP, rheumatoid factor (RF)), erosive disease and age at onset of RA.[4 16 17 23 27–30] Several prior studies have suggested that the association between HLA and erosions is due solely to CCP status; PTPN22 and CTLA4 have not been extensively studied for these phenotypes. We sought to test whether these genetic factors are associated radiographic erosions and early age of diagnosis using data from a large observational RA cohort. Thus, our goals were to (a) replicate association of *HLA, PTPN22, CTLA4*

with CCP phenotype; (b) determine if genotypes predicted radiographic erosions independent of the CCP association; and (c) to study whether age of diagnosis is associated with genetic variation. We performed conventional association testing as well as novel causal pathway analysis as proposed by Cooper[31] and Li[32] to test the hypothesis that genetic risk factors and CCP are in the causal pathway in predicting erosions.

PATIENTS AND METHODS

Study population

BRASS (Brigham Rheumatoid Arthritis Sequential Study) is a prospective observational study of 968 RA patients receiving care at the Brigham and Women's Hospital, Boston, Massachusetts, USA. The goals of the study are: (1) to determine and validate biomarkers that predict drug response and toxicity in RA, (2) to determine and validate biomarkers that predict disease activity and prognosis in RA, and (3) to evaluate the natural history of treated RA by measuring clinical, functional and economic outcomes. Baseline evaluation includes demographic and clinical information, assessment of functional status, disease activity, comorbidity, laboratory testing and hand radiographs. At a physical exam, including joint examination, assessment of pain and disease activity by medical examiner and the patient were collected at baseline and yearly. Samples of blood for immunophenoptyping, including C reactive protein (CRP), cytokines, chemokines, and rheumatoid factor (RF), anticyclic citrullinated peptide (CCP) as well as blood specimens for DNA/RNA testing were collected and stored at baseline and yearly. During follow-up, patients were mailed a self-administered questionnaire every 6 months to collect information on disease severity, functional status, resource utilisation, level of fatigue, employment status, medications, adverse events and intercurrent health events. For hand radiographs, posteroanterior (PA) views of the hands are performed at baseline and years 3 and 5 during the study. This analysis was limited to the baseline radiographs. The study was approved by the Partners Institutional Review Committee.

Laboratory methods

Blood was collected at the baseline visit for genotyping. Samples were genotyped for *HLA-SE* alleles by low-resolution genotyping, and for PTPN22, and CTLA4 (CT60 allele) by Sequenom (Sequenom, San Diego, California) genotyping. *HLA-SE* alleles 01, 04, 10, and 14 were considered positive. Rheumatoid factor testing was performed by immunoturbidimetric technique on the Cobas Integra 700 analyser (Roche Diagnostics, Indianapolis, Indiana, USA), using reagents and calibrators from Roche. Anticyclic citrullinated peptide antibodies (CCP) were measured using a second generation ELISA assay (Inova Diagnostics, Inc., San Diego, California) with a titre of >20 considered as positive. Radiograph reports were reviewed by a study rheumatologist for evidence of erosions and coded as erosion present or absent.

Statistical analysis

Association between genotypes and dichotomous phenotypes for CCP, RF and erosive RA at baseline were assessed with logistic regression models adjusting for age, sex, and disease duration. Association between genotype and age at diagnosis of RA phenotype was assessed with general linear models adjusting for sex.

Since statistical correlation does not imply a causal relationship, we performed causal pathway analysis by the method described by Li *et al*.[32] In brief, Cooper's Local Causal Discovery algorithm (Cooper's LCD) was used to explore the potential causal pathway between genotypes and phenotypes (box 1).[31] Each causal pathway analysis tested three variables: (x) genotype, (y) phenotype, and (z) phenotype. From prior knowledge we know the genotype (x), for example, HLA-SE genotype, cannot be caused by any intermediate phenotypes (y, z), for example, RF and CCP antibody status. If the pairwise unconditional correlations exists between the three variables, but the genotype (x) and one phenotype (z) is un-correlated conditional on the other phenotype (y), only one causal path can be derived from x→y→z (box 1). We conducted causal pathway analyses of the association between HLA-SE (x), CCP (y), and RF (z), and between HLA-SE (x), CCP (y), and erosion (z) phenotype.

Box 1 Criteria

- 1: x and y are correlated
- 2: y and z are correlated
- 3: x and z are **un**correlated, conditioning on y

If 1, 2 and 3 are true and since x (genotype) is not caused by either y or z, then using Cooper's LCD the only possible relationship between x, y and z is:
x→y→z
Causal pathway 1:
HLA-SE (x)→CCP (y)→RF (z)
Causal pathway 2:
HLA-SE (x)→CCP (y)→erosion (z)

RESULTS

The BRASS research study began enrollment in March 2003, and has enrolled 968 patients to date. Genotype data and autoantibody status were available for 728 subjects at the time of this analysis. For this analysis, we included only Caucasian subjects, resulting in a sample of 689 Caucasian RA patients. Among these 689 patients, mean (SD) age is 58.0 (13.8) years, mean disease duration 15.4 (12.8) years, 110 (15.9%) are recent diagnosis RA, defined as <2 years disease duration, 560 (81.2%) are female of whom 179 (32%) are in the premenopausal age

Table 1 Characteristics of 689 Caucasian subjects in the Brigham Rheumatoid Arthritis Sequential Study (BRASS) cohort genetic analyses

Parameter	Value
Mean age (SD)	58.0 (13.8)
Mean age at RA diagnosis (SD)	42.6 (15.1)
Mean years disease duration (SD)	15.4 (12.8)
Female (%)	81.2
Education (%):	
< High School	3.1
High School	18.4
Technical college, professional school	26.2
College	26.4
Graduate	25.9
Percent early onset RA (duration <2 years)	15.9
Percent CCP+	66.5
Percent RF+	61.8
Percent CCP+ and RF+	55.8
Percent with erosions	59.7

CCP, cyclic citrullinated peptide; RA, rheumatoid arthritis; RF, rheumatoid factor.

Table 2 Genotypes for *HLA-SE, PTPN22* and *CTLA 4* among 689 Caucasian rheumatoid arthritis subjects in the Brigham Rheumatoid Arthritis Sequential Study (BRASS) cohort

Gene	Genotype	n (%)
HLA-SE	None	237 (34.7)
	Single copy	272 (39.9)
	Double copy	173 (25.3)
PTPN22	CC	488 (74.4)
	CT	150 (22.9)
	TT	18 (2.7)
CTLA4	AA	116 (17.7)
	AG	320 (47.8)
	GG	220 (33.5)

range (age<51). Education level is 21.5% with high school or less and 78.5% with some college education. At baseline, 323 (47%) of patients report starting a new therapy for RA within the prior 12 months (table 1). Baseline radiographs demonstrated presence of RA erosions in 374/627 (59.7%) of subjects with radiographs. There were 61.8% RF positive subjects, and 66.5% CCP positive subjects. Mean age at diagnosis of RA was 42.58 (15.1). Genotype frequencies were similar to other RA cohorts (table 2).[4]

We first attempted to replicate the finding that alleles within three genes, *HLA-DRB1*, *PTPN22* and *CTLA4*, are associated with CCP+ RA. We created contingency tables for genotypes dichotomised by presence of any *HLA-SE* (single copy or double copy), the T allele of *PTPN22* (single or double copy), the G allele of *CTLA4* (single or double copy) and presence/absence of CCP and RF phenotypes. We tested for significance using logistic regression models assuming a multiplicative model for genotype, and adjusting for age, sex, and disease duration (table 3). Presence of any *HLA-SE* was strongly associated with CCP+ phenotype (OR 3.05, 95% CI 2.19–4.25, $p = 2.9\times10^{-9}$) and RF+ phenotype (OR 2.53, 95% CI 1.83–3.5, $p = 4.2\times10^{-7}$); presence of any PTPN22 T allele was associated with CCP+ phenotype (OR 1.81, 95% CI 1.24–2.66, $p = 0.006$) and RF+ phenotype (OR 1.84, 95% CI 1.27–2.66, $p = 0.002$); and CTLA4 was not associated with CCP (OR 1.04, 95% CI 0.7–1.55) or RF phenotypes (OR 0.94, 95% CI 0.64–1.39).

We next sought to determine whether these genetic variants were associated with two markers of disease severity, radiographic erosions and age of diagnosis. Presence of any *HLA-SE* was associated with erosive RA phenotype in unadjusted logistic regression analysis ($p = 4.0\times10^{-4}$), however this association was less strong in a logistic regression model that adjusted for age, sex and disease duration (OR 1.52, 95% CI 1.01–2.17, $p = 0.02$) (table 4). Presence of any PTPN22 T allele was not associated with erosive phenotype (OR 1.14, 95% CI 0.77–1.71), nor was CTLA4 (OR 1.34, 95% CI 0.87–2.15) (table 4).

Using general linear regression models for genotype as a predictor of age at diagnosis of RA, adjusted for sex, presence of any HLA-SE was on average associated with 3.6 years earlier age at diagnosis of RA compared with absence of HLA-SE (41.3 vs 44.9 years, $p = 0.002$) (table 5). PTPN22 was on average associated with 4.2 years earlier age at diagnosis of RA (39.5 vs 43.6 $p = 0.002$) with the earliest age at diagnosis in those with the TT genotype (37.8 years). Adjusting for sex slightly attenuated these relationships. CTLA4 genotypes were not associated with age at diagnosis of RA in this dataset.

We conducted two causal pathway analyses, adapted from Li *et al*,[32] and illustrated in box 1. We asked whether a genetic variant (HLA-SE) contributed to RF phenotype, independent of CCP status (causal pathway 1) as a replication of the Li *et al* analysis. We also asked whether HLA-SE contributed to erosion phenotype, independent of CCP status (causal pathway 2).

The first step in a causal pathway analysis[32] is to test unconditional associations between variables. We demonstrated strong relationships ($p<0.001$) between all variables (table 6).

Conditional analysis of causal pathway 1, of the association between HLA-SE (x), CCP (y), and RF (z) demonstrated that HLA-SE (x) and RF (z) are not associated when conditioning on CCP (y) (table 7). These results are similar to those shown in the causal pathway analysis by Li *et al*.[32] Therefore the evidence supports a causal pathway from *HLA-SE*→CCP→RF, but not directly from HLA-SE→RF.

Conditional analysis of causal pathway 2, of the association between *HLA-SE* (x), CCP (y), and erosion (z) demonstrated that *HLA-SE* (x) and erosion (z) are not associated when conditioning on CCP (y) (table 8). Therefore the evidence supports a causal pathway from *HLA-SE*→CCP→ erosion, but not directly from *HLA-SE*→erosion. Since the analysis presented in table 7 demonstrated that SE (x) and RF(y) are not associated when conditioning on CCP, we did not test for a causal pathway from SE→RF→erosion.

Table 3 Genotype phenotype associations with autoantibody status in the Brigham Rheumatoid Arthritis Sequential Study (BRASS) study (n = 689 Caucasian subjects)

		CCP phenotype			RF phenotype		
Gene	**Genotype**	**% CCP+**	**p Value***	**OR† (95% CI)**	**% RF+**	**p Value**	**OR (95% CI)**
HLA-SE	None	49.8	1.2×10^{-9}	2.05(1.63–2.59)	47.2	1.2×10^{-6}	1.71(1.38–2.13)
	Single copy	72.0			67.8		
	Double copy	80.6			72.2		
	Any SE	75.3	2.9×10^{-9}	3.05(2.19–4.25)	69.5	4.2×10^{-7}	2.53(1.83–3.5)
PTPN22	CC	63.4	0.008	1.67(1.14–2.43)	58.2	0.003	1.73(1.21–2.47)
	CT	75.0			71.8		
	TT	81.3			76.5		
	CT or TT	75.6	0.006	1.81(1.24–2.66)	72.3	0.002	1.84(1.27–2.66)
CTLA4	AA	63.5	0.14	1.20 (0.94–1.53)	61.7	0.48	1.09(0.86–1.37)
	AG	65.8			61.2		
	GG	69.0			62.8		
	AG or GG	67.1	0.30	1.04(0.7–1.55)	61.8	0.74	0.94(0.64–1.39)

*Logistic regression models for genotype as predictor of phenotype, adjusted for age, sex and disease duration, using a multiplicative model for genotype.
†OR is for each additional copy of *HLA-SE*, T allele of *PTPN22*, or G allele of *CTLA4*.
CCP, cyclic citrullinated peptide; OR, odds ratio; RA, rheumatoid arthritis; RF, rheumatoid factor; SE, shared epitope.

Table 4 Genotype phenotype association with erosive rheumatoid arthritis (RA) phenotype in the Brigham Rheumatoid Arthritis Sequential Study (BRASS) study (n = 689 Caucasian subjects)

		Erosion phenotype				
Gene	**Genotype**	**Erosive (%)**	**p Value***	**OR (95% CI)†**	**Adjusted p value‡**	**Adjusted OR (95% CI)**
HLA-SE	None	50	2.0×10^{-4}	1.5 (1.21–1.86)	0.005	1.38 (1.11–1.73)
	Single copy	61.9				
	Double copy	68.9				
	Any SE	64.7	4.0×10^{-4}	1.89 (1.36–2.63)	0.02	1.52 (1.01–2.17)
PTPN22	CC	58.6	0.11	1.30 (0.94–1.80)	0.2	1.26 (0.89–1.78)
	CT	59.3				
	TT	88.9				
	CT or TT	62.7	0.37	1.17 (0.82–1.67)	0.51	1.14 (0.77–1.71)
CTLA4	AA	56.0	0.81	0.97 (0.77–1.23)	0.69	1.05 (0.82–1.35)
	AG	63.3				
	GG	56.7				
	AG or GG	60.6	0.43	1.19 (0.8–1.76)	0.17	1.34 (0.87–2.15)

*Logistic regression models for genotype as predictor of phenotype, using a multiplicative model for genotype.
†OR is for each additional copy of HLA-SE, T allele of PTPN22, or G allele of CTLA4.
‡ Logistic regression models for genotype as predictor of phenotype, adjusted for age, sex and disease duration, using a multiplicative model for genotype.
OR, odds ratio; SE, shared epitope.

Table 5 RA genotypes as predictors of age at diagnosis of rheumatoid arthritis (RA) in the Brigham Rheumatoid Arthritis Sequential Study (BRASS) study (n = 689 Caucasian subjects)

Genotype	**Presence of genotype**	**Age at RA diagnosis (SD)**	**Difference***	**Unadjusted p value**	**Adjusted p value†**
HLA-SE	None	44.9 (16.7)		0.02	0.02
	Single copy	41.0 (14.7)			
	Double copy	41.9 (13.9)			
	Any SE	41.3 (14.4)	3.56	0.003	0.002
PTPN22	CC	43.6 (14.9)		0.002	0.002
	CT	39.7 (14.4)			
	TT	37.8 (16.8)			
	CT or TT	39.5 (14.7)	4.19	0.002	0.002
CTLA4	AA	42.9 (15.7)		0.72	0.64
	AG	42.1 (15.4)			
	GG	43.2 (14.0)			
	AG or GG	42.5 (15.9)	0.35	0.82	0.84

*Difference between no *HLA-SE* and double copy of *HLA-SE*, difference between CC and any T allele of *PTPN22*, and AA and any G allele of *CTLA4*.
†General linear regression model for genotype as predictor of age at diagnosis, adjusted for sex.

DISCUSSION

In this large observational RA cohort, we demonstrated strong associations between two RA genetic risk factors (*HLA-SE, PTPN22)* and RA phenotypes. *HLA-SE* was strongly associated with CCP, RF, and erosive phenotypes, even after adjusting for age, sex, and disease duration. We demonstrated that *PTPN22* was strongly associated with CCP and RF phenotypes, but not with erosive RA. We were unable to show any genotype-phenotype associations for *CTLA-4*, perhaps due to limited power as the published OR for susceptibility for *CTLA-4* is 1.2 whereas for *PTPN22* is 1.75 and for *HLA-SE* is 3.0. We found strong correlations between *HLA-SE* and the CCP and RF phenotypes, however, the *HLA-SE* association with the CCP phenotype was stronger than for RF phenotype. In this clinical cohort, using causal pathway analysis we replicated prior findings from the North American Rheumatoid Arthritis Consortium (NARAC),[18 32] and the Leiden Early Arthritis Clinic (EAC)[17] that suggest that *HLA-SE* is causally associated with the CCP phenotype, and CCP is causally associated with the RF phenotype, but *HLA-SE* is not causally associated with the RF phenotype.

Our findings for *HLA-SE* and erosions in unconditional analyses are consistent with a meta-analysis of 30 studies published from prospective cohorts and cross-sectional studies involving 3240 RA patients demonstrating an odds ratio of 2.0 (95% CI 1.8–2.2) for association of *HLA-SE* and erosions.[27] Our causal pathway analysis extends these observations to study the role of CCP antibodies. In our cross-sectional study, CCP is

Table 6 Unconditional association between variables

	χ^2	**OR (95% CI)**	**p Value**
Causal pathway 1:			
HLA-SE and CCP	44.57	3.1 (2.2–4.3)	1.0×10^{-4}
CCP and RF	278.71	24.9 (16.2–38.2)	1.0×10^{-4}
HLA-SE and RF	31.95	2.5 (1.8–3.5)	1.0×10^{-4}
Causal pathway 2:			
HLA-SE and CCP	44.57	3.1 (2.2–4.3)	1.0×10^{-4}
CCP and erosion	58.65	3.8 (2.7–5.4)	1.0×10^{-4}
HLA-SE and erosion	12.57	1.8 (1.3–2.6)	4.0×10^{-4}

CCP, cyclic citrullinated peptide; HLA-SE, Human leukocyte antigen shared epitope; RF, rheumatoid factor.

Table 7 Conditional analysis of causal pathways for HLA-SE, RF and CCP phenotypes in RA

		Conditioned on	OR_0*	OR_1†	OR_{mh}‡	p Value
x and y	SE and CCP	RF (z)	2.9 (1.6–5.2)	1.9 (1.0–3.8)	2.5 (1.6–3.8)	1.0×10^{-3}
y and z	CCP and RF	SE (x)	32.0 (15.5–66)	18.9 (10.9–32.5)	23.0 (14.9–35.5)	1.0×10^{-3}
x and z	SE and RF	CCP (y)	1.9 (0.9–3.7)	1.2 (0.7–2.1)	1.4 (0.9–2.2)	0.09

*OR_0 = Odds ratio for strata where conditioned on variable = 0 (unexposed).
†OR_1 = Odds ratio for strata where conditioned on variable = 1 (exposed).
‡OR_{mh} = Mantel-Haenszel odds ratio.
CCP, cyclic citrullinated peptide; HLA-SE, Human leukocyte antigen shared epitope; RA, rheumatoid arthritis; RF, rheumatoid factor.

Table 8 Conditional analysis of causal pathways for HLA-SE, CCP and erosion phenotypes in RA

		Conditioned on	OR_0*	OR_1†	OR_{mh}‡	p Value
x and y	SE and CCP	Erosion (z)	4.5 (2.6–7.8)	1.9 (1.1–3.3)	3.0 (2.1–4.3)	1.0×10^{-4}
y and z	CCP and erosion	SE (x)	6.0 (3.3–10.8)	2.5 (1.6–4.0)	3.5 (2.5–5.1)	1.0×10^{-4}
x and z	SE and erosion	CCP (y)	2.2 (1.2–3.9)	0.9 (0.6–1.5)	1.4 (0.9–1.9)	0.11

*OR_0 = Odds ratio for strata where conditioned on variable = 0 (unexposed).
†OR_1 = Odds ratio for strata where conditioned on variable = 1 (exposed).
‡OR_{mh} = Mantel-Haenszel odds ratio.
CCP, cyclic citrullinated peptide; HLA-SE, Human leukocyte antigen shared epitope; RA, rheumatoid arthritis; RF, rheumatoid factor.

strongly associated with erosions, but *HLA-SE* is not associated with erosions, after conditioning on CCP status, suggesting that there is no causal pathway directly from *HLA-SE* to erosions. Our causal pathway findings are consistent with two studies from the Leiden EAC.[17 30] Among RA patients in the EAC prospective cohort there were large differences in rates of erosion progression between CCP+RA and CCP-RA, with an additional effect of the *HLA-SE* on erosion progression only among the CCP+ group but no association for *HLA-SE* with erosions among the CCP– group.[17] Analysis of the undifferentiated arthritis patients in the EAC cohort followed prospectively for the development of RA demonstrated that *HLA-SE* alleles are primarily a risk factor for development of anti-CCP antibodies and are not an independent risk factor for progression to RA, after adjusting for CCP status.[30]

We found evidence of 3.6 years earlier age of diagnosis with *HLA-SE* in this large clinical cohort with a mean disease duration of 15 years. This work replicates other studies in which *HLA-SE* was associated with 6 years earlier age at onset in a seropositive RA cohort with <15 months of disease duration in the US[28] as well as in a Korean population in which the specific allele, *HLA-DRB1**0405, was associated with 4 years earlier onset.[33] We found evidence for 4.2 years earlier diagnosis of RA for *PTPN22,* which is similar to the findings of 2 years earlier age of onset in samples from North America and Sweden.[4] In a population from the UK, *PTPN22* was associated with 8.6 years earlier onset in homozygotes, and 4.7 years earlier onset in heterozygotes.[34]

The BRASS cohort is a well-educated, primarily Caucasian population with long disease duration, treated at a tertiary referral centre in the United States, all factors that may limit generalisability. Although disease duration is similar to that in the North American Rheumatoid Arthritis Consortium (NARAC), the rates of seropositivity and erosive disease are lower, since by design NARAC recruited more severe RA patients.[32] However age, disease duration, rates of erosive disease, and seropositivity are similar to those reported in the National Databank study of >14 000 patients enrolled from rheumatology practices across the US,[35 36] suggesting that BRASS subjects are more similar to RA patients seen in the community. The causal pathway approach is a statistical method that requires a number of assumptions, as discussed in detail by Cooper *et al*,[31] and it is possible that our dataset does not meet all of the assumptions. The approach does allow for the presence of potential confounders, as long as the variable "x", in this case genotype, is not caused by the confounder.

In conclusion, genotype-phenotype analysis of a large clinical cohort demonstrates the importance of considering phenotypes when studying genetic predictors. We replicated association of *HLA-SE*, and *PTPN22* with CCP+ RA compared with CCP– RA as well as associations with earlier diagnosis of RA, but were unable to demonstrate any associations for *CTLA-4*. Of the three genes studied, only *HLA-SE* was associated with radiographic erosions but this association was not independent of CCP status. The novel causal pathway analyses confirms prior studies that demonstrate the importance of antibodies to CCP in the pathogenesis of joint damage in RA and provides support to recent calls[17 23] for considering CCP+ RA as a separate clinical entity within the overall RA phenotype.

Acknowledgements: We wish to thank the BRASS participants, the rheumatologists in the Robert B Brigham Arthritis Center for their extensive time and efforts on behalf of the study, and our talented team of research assistants.

Funding: EWK is supported by NIH grants R01 AR49880, and K24 AR0524-01. RMP is supported by NIH K08 A155314. The BRASS cohort and co-authors JC, RJG, NEM, EI, MEW, and NAS are supported by Millenium Pharmaceuticals. AP was formerly supported by Millenium Pharmaceuticals and is currently supported by Amgen. RR was formerly supported by Millenium Pharmaceuticals and is currently supported by Biogen/IDEC.

Competing interests: None declared.

REFERENCES

1. **Gregersen PK,** Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987;**30**:1205–13.
2. **MacGregor AJ,** Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, *et al*. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000;**43**:30–7.
3. **Begovich AB,** Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, *et al*. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004;**75**:330–7.
4. **Plenge RM,** Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, *et al*. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. *Am J Hum Genet* 2005;**77**:1044–60.
5. **Jawaheer D,** Gregersen PK. Rheumatoid arthritis. The genetic components. *Rheum Dis Clin North Am* 2002;**28**:1–15.

6. **Newton JL,** Harney SM, Wordsworth BP, Brown MA. A review of the MHC genetics of rheumatoid arthritis. *Genes Immun* 2004;**5**:151–7.
7. **Jawaheer D,** Seldin MF, Amos CI, Chen WV, Shigeta R, Etzel C, *et al*. Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. *Arthritis Rheum* 2003;**48**:906–16.
8. **Lee AT,** Li W, Liew A, Bombardier C, Weisman M, Massarotti EM, *et al*. The PTPN22 R620W polymorphism associates with RF positive rheumatoid arthritis in a dose-dependent manner but not with HLA-SE status. *Genes Immun* 2005;**6**:129–33.
9. **Hinks A,** Barton A, John S, Bruce I, Hawkins C, Griffiths CE, *et al*. Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: further support that PTPN22 is an autoimmunity gene. *Arthritis Rheum* 2005;**52**:1694–9.
10. **Orozco G,** Sanchez E, Gonzalez-Gay MA, Lopez-Nevot MA, Torres B, Caliz R, *et al*. Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Rheum* 2005;**52**:219–24.
11. **Prescott NJ,** Fisher SA, Onnie C, Pattni R, Steer S, Sanderson J, *et al*. A general autoimmunity gene (PTPN22) is not associated with inflammatory bowel disease in a British population. *Tissue Antigens* 2005;**66**:318–20.
12. **Viken MK,** Amundsen SS, Kvien TK, Boberg KM, Gilboe IM, Lilleby V, *et al*. Association analysis of the 1858C>T polymorphism in the PTPN22 gene in juvenile idiopathic arthritis and other autoimmune diseases. *Genes Immun* 2005;**6**:271–3.
13. **Zhernakova A,** Eerligh P, Wijmenga C, Barrera P, Roep BO, Koeleman BP. Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. *Genes Immun* 2005;**6**:459–61.
14. **Criswell LA,** Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, *et al*. Analysis of families in the Multiple Autoimmune Disease Genetics Consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am J Hum Genet* 2005;**76**:561–71.
15. **Lee YH,** Rho YH, Choi SJ, Ji JD, Song GG, Nath SK, *et al*. The PTPN22 C1858T functional polymorphism and autoimmune diseases – a meta-analysis. *Rheumatology (Oxford)* 2007;**46**:49–56.
16. **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* 2004;**50**:3085–92.
17. **Huizinga TW,** Amos CI, van der Helm-van Mil AH, Chen W, van Gaalen FA, Jawaheer D, *et al*. Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. *Arthritis Rheum* 2005;**52**:3433–8.
18. **Irigoyen P,** Lee AT, Wener MH, Li W, Kern M, Batliwalla F, *et al*. Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles. *Arthritis Rheum* 2005;**52**:3813–8.
19. **Johansson M,** Arlestig L, Hallmans G, Rantapaa-Dahlqvist S. PTPN22 polymorphism and anti-cyclic citrullinated peptide antibodies in combination strongly predicts future onset of rheumatoid arthritis and has a specificity of 100% for the disease. *Arthritis Res Ther* 2006;**8**:R19.
20. **Vaidya B,** Pearce SH, Charlton S, Marshall N, Rowan AD, Griffiths ID, *et al*. An association between the CTLA4 exon 1 polymorphism and early rheumatoid arthritis with autoimmune endocrinopathies. *Rheumatology (Oxford)* 2002;**41**:180–3.
21. **Rodriguez MR,** Nunez-Roldan A, Aguilar F, Valenzuela A, Garcia A, Gonzalez-Escribano MF. Association of the CTLA4 3′ untranslated region polymorphism with the susceptibility to rheumatoid arthritis. *Hum Immunol* 2002;**63**:76–81.
22. **Suppiah V,** O'Doherty C, Heggarty S, Patterson CC, Rooney M, Vandenbroeck K. The CTLA4+49A/G and CT60 polymorphisms and chronic inflammatory arthropathies in Northern Ireland. *Exp Mol Pathol* 2006;**80**:141–6.
23. **Klareskog L,** Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, *et al*. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum* 2006;**54**:38–46.
24. **Linn-Rasker SP,** van der Helm-van Mil AH, van Gaalen FA, Kloppenburg M, de Vries RR, le Cessie S, *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* 2006;**65**:366–71.
25. **Pedersen M,** Jacobsen S, Klarlund M, Pedersen BV, Wiik A, Wohlfahrt J, *et al*. Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. *Arthritis Res Ther* 2006;**8**:R133.
26. **Pedersen M,** Jacobsen S, Garred P, Madsen HO, Klarlund M, Svejgaard A, *et al*. Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control study in Denmark. *Arthritis Rheum* 2007;**56**:1446–53.
27. **Gorman JD,** Lum RF, Chen JJ, Suarez-Almazor ME, Thomson G, Criswell LA. Impact of shared epitope genotype and ethnicity on erosive disease: a meta-analysis of 3,240 rheumatoid arthritis patients. *Arthritis Rheum* 2004;**50**:400–12.
28. **Wu H,** Khanna D, Park G, Gersuk V, Nepom GT, Wong WK, *et al*. Interaction between RANKL and HLA-DRB1 genotypes may contribute to younger age at onset of seropositive rheumatoid arthritis in an inception cohort. *Arthritis Rheum* 2004;**50**:3093–103.
29. **van Gaalen FA,** van Aken J, Huizinga TW, Schreuder GM, Breedveld FC, Zanelli E, *et al*. Association between HLA class II genes and autoantibodies to cyclic citrullinated peptides (CCPs) influences the severity of rheumatoid arthritis. *Arthritis Rheum* 2004;**50**:2113–21.
30. **van der Helm-van Mil AH,** Verpoort KN, Breedveld FC, Huizinga TW, Toes RE, de Vries RR. The HLA-DRB1 shared epitope alleles are primarily a risk factor for anti-cyclic citrullinated peptide antibodies and are not an independent risk factor for development of rheumatoid arthritis. *Arthritis Rheum* 2006;**54**:1117–21.
31. **Cooper GF.** A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Mining Knowledge Disc* 1997;**1**:203–24.
32. **Li W,** Wang M, Irigoyen P, Gregersen PK. Inferring causal relationships among intermediate phenotypes and biomarkers: a case study of rheumatoid arthritis. *Bioinformatics* 2006;**22**:1503–7.
33. **Lee HS,** Lee KW, Song GG, Kim HA, Kim SY, Bae SC. Increased susceptibility to rheumatoid arthritis in Koreans heterozygous for HLA-DRB1*0405 and *0901. *Arthritis Rheum* 2004;**50**:3468–75.
34. **Steer S,** Lad B, Grumley JA, Kingsley GH, Fisher SA. Association of R602W in a protein tyrosine phosphatase gene with a high risk of rheumatoid arthritis in a British population: evidence for an early onset/disease severity effect. *Arthritis Rheum* 2005;**52**:358–60.
35. **Choi HK,** Hernan MA, Seeger JD, Robins JM, Wolfe F. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. *Lancet* 2002;**359**:1173–7.
36. **Finckh A,** Choi HK, Wolfe F. Progression of radiographic joint damage in different eras: trends towards milder disease in rheumatoid arthritis are attributable to improved treatment. *Ann Rheum Dis* 2006;**65**:1192–7.