Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility

Anne Barton[1],[*], Wendy Thomson[1], Xiayi Ke[1], Steve Eyre[1], Anne Hinks[1], John Bowes[1], Laura Gibbons[1], Darren Plant[1], Wellcome Trust Case Control Consortium[7], Anthony G. Wilson[2], Ioanna Marinou[2], Ann Morgan[3], Paul Emery[3], YEAR consortium[8], Sophia Steer[4], Lynne Hocking[5], David M. Reid[5], Paul Wordsworth[6], Pille Harrison[6] and Jane Worthington[1]

[1]Arc-Epidemiology Unit, Stopford Building, The University of Manchester, Manchester, UK, [2]School of Medicine and Biomedical Sciences, The University of Sheffield, Sheffield S10 2JF, UK, [3]Academic Unit of Musculoskeletal Disease, Chapel Allerton Hospital, Leeds LS7 4SA, UK, [4]Clinical and Academic Rheumatology, Kings College Hospital NHS Foundation Trust, Denmark Hill, London SE5 9RS, UK, [5]Bone Research Group, Department of Medicine and Therapeutics, University of Aberdeen, Aberdeen AB25 2ZD, UK, [6]University of Oxford Institute of Musculoskeletal Sciences, Botnar Research Centre, Oxford OX3 7LD, UK, [7]WTCCC (Supplementary Note online) and [8]YEAR Consortium (Supplementary Note online)

Rheumatoid arthritis (RA) is an archetypal, common, complex autoimmune disease with both genetic and environmental contributions to disease aetiology. Two novel RA susceptibility loci have been reported from recent genome-wide and candidate gene association studies. We, therefore, investigated the evidence for association of the *STAT4* and *TRAF1/C5* loci with RA using imputed data from the Wellcome Trust Case Control Consortium (WTCCC). No evidence for association of variants mapping to the *TRAF1/C5* gene was detected in the 1860 RA cases and 2930 control samples tested in that study. Variants mapping to the *STAT4* gene did show evidence for association (rs7574865, *P* = 0.04). Given the association of the *TRAF1/C5* locus in two previous large case–control series from populations of European descent and the evidence for association of the *STAT4* locus in the WTCCC study, single nucleotide polymorphisms mapping to these loci were tested for association with RA in an independent UK series comprising DNA from >3000 cases with disease and >3000 controls and a combined analysis including the WTCCC data was undertaken. We confirm association of the *STAT4* and the *TRAF1/C5* loci with RA bringing to 5 the number of confirmed susceptibility loci. The effect sizes are less than those reported previously but are likely to be a more accurate reflection of the true effect size given the larger size of the cohort investigated in the current study.

Rheumatoid arthritis [RA (MIM 180300)] is a chronic inflammatory arthritis occurring in 0.8% UK population and characterized by progressive destruction of synovial joints (1). A strong genetic component to disease aetiology has been determined with heritability estimates of 50–60% (2).

The major susceptibility loci are: (i) the *HLA-DRB1* gene, in which a group of alleles each sharing a common amino acid sequence in the peptide-binding groove and collectively referred to as the shared epitope, is associated with both susceptibility and severity to RA (3) and (ii) the protein tyrosine

---

Table 1. Imputed genotypes for RA cases and controls from WTCCC study

Locus	RA cases	Population controls	OR (95% CI)	Trend P
rs3761847 (*TRAF1/C5*) G/A	1/1 608 (32.7)	1/1 960 (32.7)	0.99 (0.91–1.07)	0.80
	1/2 902 (48.5)	1/2 1410 (48.0)		
	2/2 348 (18.7)	2/2 565 (19.3)		
rs10818488 (*TRAF1/C5*) A/G	1/1 578 (31.1)	1/1 920 (31.4)	1.00 (0.92–1.09)	0.93
	1/2 911 (49.0)	1/2 1419 (48.4)		
	2/2 369 (19.9)	2/2 595 (20.3)		
rs10760130 (*TRAF1/C5*) G/A	1/1 578 (31.1)	1/1 920 (31.4)	1.00 (0.92–1.09)	0.93
	1/2 911 (49.0)	1/2 1419 (48.4)		
	2/2 369 (19.9)	2/2 595 (20.3)		
rs10760129 (*TRAF1/C5*) T/C	1/1 578 (31.1)	1/1 920 (31.4)	1.00 (0.92–1.09)	0.93
	1/2 911 (49.0)	1/2 1419 (48.4)		
	2/2 369 (19.9)	2/2 595 (20.3)		
rs10181656 (*STAT4*) C/G	1/1 1050 (56.5)	1/1 1736 (59.1)	1.11 (1.01–1.22)	0.03
	1/2 692 (37.2)	1/2 1047 (35.7)		
	2/2 117 (6.3)	2/2 151 (5.2)		
rs8179673 (*STAT4*) T/C	1/1 1054 (56.7)	1/1 1742 (59.4)	1.11 (1.01–1.22)	0.04
	1/2 689 (37.1)	1/2 1042 (35.5)		
	2/2 115 (6.2)	2/2 150 (5.1)		
rs11889341 (*STAT4*) C/T	1/1 1092 (58.8)	1/1 1800 (61.3)	1.10 (1.00–1.22)	0.05
	1/2 671 (36.1)	1/2 1010 (34.4)		
	2/2 95 (5.1)	2/2 125 (4.3)		
rs7574865 (*STAT4*) G/T	1/1 1080 (58.1)	1/1 1782 (60.7)	1.11 (1.00–1.22)	0.04
	1/2 675 (36.3)	1/2 1016 (34.6)		
	2/2 103 (5.5)	2/2 136 (4.6)		

Data shown in n (%).
1, common allele, listed first; 2, minor allele, listed second.

phosphatase 22 (*PTPN22*) gene, which is associated with susceptibility to RA (4). Together, they account for 50% of the genetic susceptibility to disease in populations of European descent, although, interestingly, variation across the *PTPN22* gene does not appear to be associated with RA in Japanese or Korean populations (5).

Recently, tremendous progress has been made in identifying further RA susceptibility variants and this has been achieved using both genome-wide association (GWA) and candidate gene approaches. First, the Wellcome Trust Case Control Consortium (WTCCC) study was a GWA that included 1860 RA cases and 2930 controls and confirmed association to the two known susceptibility variants, *HLA-DRB1* and *PTPN22* ($P < 10^{-7}$) (6). In addition, nine other loci showed modest evidence for significance and association to one of these, a locus lying between the *OLIG3* and *TNFAIP3* genes on chromosome 6q, has been unequivocally replicated in UK and US populations (7,8). Secondly, a GWA study in US and Swedish populations reported association to the *TRAF1/C5* locus and this has subsequently been confirmed by a separate study (9,10). Finally, the results of a fine-mapping strategy investigating candidate genes mapping under a peak of linkage in US RA families has identified another RA susceptibility locus mapping to the *STAT4* gene in US subjects, which has been subsequently confirmed in Swedish and Korean populations (11,12).

Many of the variants putatively associated with RA susceptibility were not genotyped directly in the WTCCC study. However, genotypes have been imputed using the data from directly genotyped single nucleotide polymorphisms (SNPs) nearby and from patterns of linkage disequilibrium inferred from HapMap data. Hence, imputed genotypes are available from the WTCCC study for all common HapMap SNPs and

this data provides an opportunity to explore the two candidate RA susceptibility loci, *STAT4* and *TRAF1/C5*, in this data set.

The aim of the current study was to look for evidence for association of previously reported RA susceptibility loci in the WTCCC data set and to validate the findings by direct genotyping of SNPs mapping to these regions in an independent cohort of 3418 cases and 3337 controls from the UK.

RESULTS

Imputed data

Imputed genotype frequencies conformed to Hardy–Weinberg expectations in the control population and allele frequencies were similar to those reported previously in European populations where data were available.

No difference in genotype frequencies was observed for the SNPs mapping to the *TRAF1/C5* locus, previously reported to be associated with RA in other populations, in the UK WTCCC series (Table 1). Nominal evidence for association was observed for SNPs mapping to the *STAT4* gene (Table 1).

Validation

The same SNPs were genotyped directly in an independent series of 3418 RA cases and 3337 controls. Concordance rate for duplicate samples was 99.5%. A Breslow Day test was undertaken to investigate heterogeneity between the samples recruited by the different centres but none was observed for any of the SNPs ($P > 0.1$). Hence genotype counts were combined across the centres and compared between validation cases and validation controls. SNPs

Table 2. Validation of SNPs in independent data set and combined analysis (with WTCCC samples)

SNP	Minor allele[a] and frequency	Validation study Genotypes (cases versus controls)	OR (95% CI)	Trend-*P*	Combined analysis Genotypes (cases versus controls)	OR (95% CI)	Trend-*P*
rs3761847 (*TRAF1/C5*)	A and 0.43	1073 (31.4) versus 1102 (33.0) 1612 (47.2) versus 1607 (48.2) 733 (21.4) versus 628 (18.8)	1.09 (1.02–1.17)	0.014	1681 (31.9) versus 2060 (32.9) 2514 (47.6) versus 3017 (48.1) 1081 (20.5) versus 1193 (19.0)	1.08 (1.00–1.11)	0.06
rs10818488 (*TRAF1/C5*)	G and 0.43	988 (29.0) versus 968 (31.9) 1687 (49.5) versus 1514 (49.9) 734 (21.5) versus 552 (18.2)	1.14 (1.06–1.22)	3.5×10^{-4}	1566 (29.7) versus 1888 (31.6) 2598 (49.3) versus 2933 (49.1) 1103 (20.9) versus 1147 (19.2)	1.08 (1.02–1.13)	0.006
rs10760130 (*TRAF1/C5*)	A and 0.44	949 (27.9) versus 943 (31.1) 1693 (49.7) versus 1524 (50.2) 764 (22.4) versus 567 (18.7)	1.15 (1.07–1.23)	7.1×10^{-5}	1527 (29.0) versus 1863 (31.2) 2604 (49.5) versus 2943 (49.3) 1133 (21.5) versus 1162 (19.5)	1.09 (1.03–1.15)	0.001
rs10760129 (*TRAF1/C5*)	C 0.44	957 (28.1) versus 946 (31.3) 1682 (49.4) versus 1516 (50.1) 765 (22.5) versus 563 (18.6)	1.15 (1.08–1.24)	6.5×10^{-5}	1535 (29.2) versus 1866 (31.3) 2593 (49.2) versus 2935 (49.3) 1134 (21.5) versus 1158 (19.4)	1.09 (1.03–1.15)	0.002
rs10181656 (*STAT4*)	G and 0.22	1958 (57.3) versus 2051 (61.5) 1264 (37.0) versus 1139 (34.1) 196 (5.7) versus 147 (4.4)	1.17 (1.08–1.27)	1.3×10^{-4}	3008 (57.0) versus 3787 (60.4) 1956 (37.1) versus 2186 (34.9) 313 (5.9) versus 298 (4.8)	1.14 (1.07–1.21)	4.0×10^{-5}
rs8179673 (*STAT4*)	C and 0.22	1959 (57.3) versus 2051 (61.5) 1262 (36.9) versus 1140 (34.2) 197 (5.8) versus 146 (4.4)	1.17 (1.08–1.27)	1.2×10^{-4}	3013 (57.1) versus 3793 (60.5) 1951 (37.0) versus 2182 (34.8) 312 (5.9) versus 296 (4.7)	1.14 (1.07–1.21)	3.9×10^{-5}
rs11889341 and (*STAT4*)	T and 0.22	1969 (57.6) versus 2057 (61.6) 1257 (36.8) versus 1134 (34.2) 192 (5.6) versus 146 (4.4)	1.16 (1.07–1.26)	2.4×10^{-4}	3061 (58.0) versus 3857 (61.5) 1928 (36.5) versus 2144 (34.2) 287 (5.4) versus 271 (4.3)	1.14 (1.07–1.21)	2.7×10^{-5}
rs7574865 (*STAT4*)[b]	T and 0.22	1937 (57.0) versus 1860 (61.5) 1266 (37.2) versus 1024 (33.9) 196 (5.8) versus 140 (4.6)	1.17 (1.08–1.28)	1.4×10^{-4}	3017 (57.4) versus 3642 (61.1) 1941 (36.9) versus 2040 (34.2) 299 (5.7) versus 276 (4.6)	1.15 (1.08–1.22)	1.9×10^{-5}

[a]Allele assignments as per Ensembl (www.ensembl.org) release 49, March 2008, from sense strand.
[b]Only genotyped in 3399 cases and 3024 controls.

mapping to both the *STAT4* and *TRAF1/C5* loci were significantly associated with RA susceptibility in the validation cohort (Table 2), although effect sizes were lower than had been reported in the previous, smaller studies.

Combined analysis

The imputed data from the WTCCC study were firstly combined with those from the validation study to create a combined sample of >5000 RA cases and ~6000 controls in order to allow robust estimates of the strengths of the

effect sizes for these two loci to be calculated. The four SNPs mapping to each locus show strong correlation with each other (r^2 between *STAT4* SNPs = >0.97; r^2 between *TRAF1/C5* SNPs = 0.97). The largest effect size of the SNPs mapping to the *STAT4* locus arises from rs7574865 (OR 1.15, 95% CI 1.08, 1.22; $P = 1.9 \times 10^{-5}$) and this is in line with previous studies. Of the four *TRAF1/C5* SNPs tested, rs10760130 showed the greatest statistical evidence for association but the effect size was less than that of the *STAT4* locus (OR 1.09, 95% CI 1.03, 1.15; $P = 0.001$). Secondly, a meta-analysis including both the WTCCC and vali-

Table 3. Stratification by CCP-positive/negative subgroups in validation data set (controls are the same as shown in Table 2)

SNP	Genotypes (CCP+)	Genotypes (CCP−)	CCP+ versus controls[a] OR (95% CI)	Trend-*P*	CCP− versus controls[a] OR (95% CI)	Trend-*P*
rs3761847 (*TRAF1/C5*)	257 (32.4)	166 (34.2)	1.09 (0.97–1.21)	0.14	1.05 (0.91–1.20)	0.51
	359 (45.3)	212 (43.6)				
	177 (22.3)	108 (22.2)				
rs10818488 (*TRAF1/C5*)	359 (29.6)	190 (30.8)	1.10 (1.00–1.21)	0.046	1.08 (0.95–1.22)	0.23
	605 (49.8)	299 (48.5)				
	250 (20.6)	128 (20.7)				
rs10760130 (*TRAF1/C5*)	342 (28.2)	185 (30.0)	1.12 (1.02–1.23)	0.018	1.08 (0.96–1.22)	0.22
	609 (50.3)	300 (48.6)				
	260 (21.5)	132 (21.4)				
rs10760129 (*TRAF1/C5*)	346 (28.5)	186 (30.1)	1.12 (1.02–1.23)	0.017	1.08 (0.96–1.22)	0.20
	606 (50.0)	299 (48.5)				
	261 (21.5)	132 (21.4)				
rs10181656 (*STAT4*)	451 (56.9)	279 (57.4)	1.15 (1.01–1.31)	0.03	1.21 (1.03–1.41)	0.02
	305 (38.5)	173 (35.6)				
	37 (4.7)	34 (7.0)				
rs8179673 (*STAT4*)	452 (57.0)	279 (57.4)	1.15 (1.01–1.30)	0.04	1.21 (1.03–1.41)	0.02
	304 (38.3)	173 (35.6)				
	37 (4.7)	34 (7.0)				
rs11889341 (*STAT4*)	451 (56.9)	280 (57.6)	1.15 (1.01–1.31)	0.04	1.20 (1.03–1.40)	0.02
	307 (38.7)	173 (35.6)				
	35 (4.4)	33 (6.8)				
rs7574865 (*STAT4*)	674 (55.7)	347 (56.2)	1.20 (1.07–1.34)	0.0014	1.21 (1.05–1.40)	0.0088
	475 (39.2)	232 (37.6)				
	62 (5.1)	38 (6.2)				

[a]A direct comparison of CCP+ versus CCP− samples showed that there was no significant difference between these two groups.

dation studies was undertaken, which yielded very similar odds ratios (OR rs7574865 1.14, 95% CI 1.07, 1.22; OR rs10760130 1.08, 95% CI 1.03, 1.14).

Stratification by anti-cyclic-citrullinated peptide antibody status

RA is a clinically heterogeneous disease and there has been some speculation recently that it may comprise two distinct subgroups characterized by the presence or absence of antibodies to citrullinated peptides recognized by anti-cyclic-citrullinated peptide (CCP) antibodies. For example, the association of carriage of shared epitope alleles of the *HLA-DRB1* gene, the major RA susceptibility gene, appears confined to anti-CCP-positive RA cases (reviewed in 13). Hence, subgroup analysis was undertaken to explore whether the associations of the *STAT4* and *TRAF1/C5* loci differed in anti-CCP-positive and -negative individuals (Table 3). For the *STAT4* locus, the strength of the association was similar in both subgroups. Although the *TRAF1/C5* associations appeared stronger in the anti-CCP-positive subgroup, an effect, albeit smaller, was observed in the negative subgroup and a direct comparison between anti-CCP-positive and -negative subjects showed no statistically significant evidence for a difference between the subgroups (Table 3).

DISCUSSION

We have performed the best powered study, to date, investigating putative RA susceptibility variants and have confirmed that both *STAT4* and the *TRAF1/C5* loci are associated with RA susceptibility. This brings to 5 the number of confirmed RA susceptibility genes (*HLA DRB1, PTPN22* and *OLIG3/ TNFAIP3* in addition to the above two). The large sample size allows robust estimates of the effect sizes to be inferred and suggests that the remaining RA susceptibility genes may have smaller effect sizes.

It is interesting to note that the *TRAF1/C5* locus showed no evidence for association in the WTCCC series but significant evidence for association in the larger validation cohort, despite the similarity in allele and genotype frequencies between the two UK control groups analysed. The linkage disequilibrium across the region is high and confidence scores for the imputed genotypes exceeded 97%, suggesting that the imputed genotypes from the WTCCC study should be accurate. Indeed, although none of the previously associated markers was genotyped directly in the WTCCC study, a SNP in near-perfect correlation (rs10118357 $r^2 = 0.97$ with rs3761847 and $r^2 = 1$ with rs10818488) was genotyped and also failed to show evidence for association with RA in that series (rs10118357 $P = 0.89$).

The different effect sizes observed at the *TRAF1/C5* locus in different populations could be explained by population heterogeneity but we feel that this is unlikely because first, clinical, demographic and control allele frequencies were similar in the two UK cohorts investigated, only one of which showed evidence for association to the locus. Secondly, clinical and demographic characteristics of the UK RA cases were similar to the US, Swedish and Dutch series and, finally, the frequency of other RA associated variants is similar across all the populations studied. For example, published data on

the rs2476601 SNP in the *PTPN22* gene suggests that effect sizes are similar across the two UK cohorts, the US and Dutch series (6,7,14,15).

A more likely explanation of the differences of effect seen across the different cohorts is that the effect size at this locus is very modest, and it is, therefore, not surprising that the SNPs will be associated in some, but not all, studies. Surprisingly, this locus was first identified from two independent studies in which the initial sample sizes were much smaller than those tested in the WTCCC study (9,10). This aptly demonstrates the phenomenon of 'winner's curse' in which effect sizes are often over-estimated in the original studies. It also demonstrates that failure to detect evidence for association, even in a large case–control series like the WTCCC sample, does not exclude a true but modest effect being present. Indeed, if the true effect size is more like 1.09 and given the risk allele frequency, the WTCCC study would only have had 54% power to detect it at the 5% significance level.

The evidence for association of SNPs mapping to the *STAT4* gene is more consistent, in keeping with the larger effect size conferred by this locus. The locus has previously been associated with both RA and systemic lupus erythematosus (SLE) (11), suggesting that it may predispose to a number of auto-immune diseases in the same way as the *PTPN22* functional variant has been shown to. In fact, the strength of the association may be more pronounced in SLE. Further exploration in other autoimmune diseases is now warranted.

Sequencing and fine mapping studies will be required to identify the most likely causal polymorphisms in these loci. However, the strong genetic correlation of the SNPs in both the *STAT4* and *TRAF1/C5* regions mean that genetic studies alone will be unlikely to identify the causal variant. Functional studies will be necessary to determine which the disease variant is and how it contributes to the development of RA but already it is possible to hypothesize a potential pathway: previous evidence suggests that carriage of the *PTPN22* susceptibility variant leads to a failure to delete autoreactive T-cells, thereby predisposing to autoimmunity in general (16). The presence of cigarette smoke, a recognized environmental risk factor for RA, has been shown to result in the citrullination of proteins (17) and the shared epitope alleles of the *HLA DRB1* gene are known to bind citrullinated epitopes more efficiently resulting in an exaggerated helper T-cell response (18). Variation in either gene could predispose to RA, therefore, by creating a permissive environment for an up-regulated immune response. In turn, IL-12-induced activation of STAT4 has been shown to drive the production of Th1 and Th17 rather than Th2 cells, and a Th1/Th17 mediated inflammatory response is a hallmark of RA (19). The exaggerated Th1/Th17 response may mean that, in the presence of an inflammatory trigger, the inflammatory response is up-regulated leading to excess tumour necrosis factor (TNF) production with resultant binding of TNF to its cell-surface receptors TNFR1 and 2. Signalling via these receptors is mediated via a number of pathways but the TRAF family of receptors has been implicated, including TRAF1. In turn, TRAF1 signalling has been reported to activate TNFAIP3 (A20)-induced apoptosis and activation of NF-κB (20). Variation in the genes or regulatory sequences at any stage of this pathway could, therefore, predispose to RA by producing

a prolonged and sustained inflammatory response characteristic of RA. Functional studies will be required to confirm or refine this hypothesis and determine where environmental susceptibility factors interact.

The studies confirm the utility of GWA approaches and well-powered studies in unravelling the intricacies of common, complex diseases such as RA. It is likely that a number of other susceptibility loci of more moderate effect sizes exist and these may only be revealed through combining data or meta-analysis to generate sufficiently large sample sizes to be confident of identifying their effects. However, already key pathways are emerging and provide the possibility of the development of further targeted therapies for this disabling condition in future.

METHODS

Study design

Imputed genotypes for SNP markers previously reported to be associated with RA were compared between 1860 RA cases and 2930 controls from the WTCCC study. Previous work suggests that, where linkage disequilibrium is high and confidence scores for imputed genotypes exceed 95%, the accuracy of imputation in predicting actual genotype counts exceeds 98.4% (21). Variants mapping to the *STAT4* and *TRAF1/C5* genes previously reported to be associated with RA in other populations, were genotyped in an independent UK sample of 3418 cases with RA and 3337 controls. A combined analysis was then undertaken including all the data from the WTCCC and the validation cohort to provide a robust estimate of effect sizes for these variants.

Cohorts

All subjects were white Caucasians and all cases satisfied 1987 ACR classification criteria for RA, modified for genetic studies. Both cohorts have been described previously but briefly:

WTCCC cohort. Cases ($n = 1860$) with RA were recruited as part of the **arc** National Repository of Family Material, from a primary care-based inception cohort or from NHS Rheumatology Clinics throughout the UK. Control data ($n = 2930$) were available for the 1958 birth cohort and from blood donors as described and reported previously (6).

Validation data set. DNA was available from RA patients recruited from six centres in the UK and five of these centres also provided DNA samples from healthy controls: Manchester 1372 cases and 924 controls (including 357 controls from the 1958 birth cohort, not overlapping with those samples tested in the WTCCC study); Sheffield 979 cases and 995 controls; Leeds, 1126 cases and 532 controls; Aberdeen, 523 cases and 862 controls; Oxford, 736 cases and 536 controls; London, 327 cases, as reported previously. The cases were from NHS Rheumatology Clinics throughout the UK. Presence of autoantibodies [rheumatoid factor (RF) and/or anti-CCP] was documented for a proportion of the patients.

SNP selection

Imputed genotype data for the following SNPs were retrieved from the WTCCC Web site:

STAT4: rs7574865, rs10181656, rs8179673 and rs11889341 show high correlation (r^2) with each other, map to the *STAT4* gene and have been associated following a fine mapping study of a peak of linkage in US samples and replicated in Swedish and Korean cohorts (11,12);

TRAF1/C5: rs3761847, rs10818488, rs10760130, rs10760129 and rs2900180 map to the *TRAF1/C5* locus and these SNPs (or highly correlated proxies) have been associated with RA in populations from the Netherlands, US and Sweden (9,10).

Genotyping and analysis

Genotyping of the same variants in an independent cohort of RA cases and controls was undertaken using the Sequenom MassArray platform (Sequenom, Inc., San Diego, CA, USA). Negative controls and duplicate samples were included to ensure the accuracy of genotyping. Only those samples with a >90% success rate and only those SNPs with a genotyping success rate of >95% were included in the analysis. Allele and genotype frequencies were compared between RA patients and population controls using the trend test implemented in STATA. Combined analysis of SNPs using imputed data from the WTCCC study and the validation cohort was subsequently performed and, in this combined group, stratification analysis based on the presence of anti-CCP antibodies was undertaken.

This study was approved by the North West Multicentre Research Ethics Committee (MREC 99/8/84) and all subjects provided informed consent.

REFERENCES

1. Symmons, D., Turner, G., Webb, R., Asten, P., Barrett, E., Lunt, M., Scott, D. and Silman, A. (2002) The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. *Rheumatology (Oxford)*, **41**, 793–800.
2. MacGregor, A.J., Snieder, H., Rigby, A.S., Koskenvuo, M., Kaprio, J., Aho, K. and Silman, A.J. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.*, **43**, 30–37.
3. Gregersen, P.K., Silver, J. and Winchester, R.J. (1987) The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum.*, **30**, 1205–1213.
4. Hinks, A., Worthington, J. and Thomson, W. (2006) The association of PTPN22 with rheumatoid arthritis and juvenile idiopathic arthritis. *Rheumatology (Oxford)*, **45**, 365–368.
5. Ikari, K., Momohara, S., Inoue, E., Tomatsu, T., Hara, M., Yamanaka, H. and Kamatani, N. (2006) Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. *Rheumatology (Oxford)*, **45**, 1345–1348.
6. Welcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature*, **447**, 661–678.
7. Thomson, W., Barton, A., Ke, X., Eyre, S., Hinks, A., Bowes, J., Donn, R., Symmons, D., Hider, S., Bruce, I.N. *et al.* (2007) Rheumatoid arthritis association at 6q23. *Nat. Genet.*, **39**, 1431–1433.
8. Plenge, R.M., Cotsapas, C., Davies, L., Price, A.L., de Bakker, P.I., Maller, J., Pe'er, I., Burtt, N.P., Blumenstiel, B., Defelice, M. *et al.* (2007) Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.*, **39**, 1477–1482.
9. Kurreeman, F.A., Padyukov, L., Marques, R.B., Schrodi, S.J., Seddighzadeh, M., Stoeken-Rijsbergen, G., van der Helm-van Mil, A.H., Allaart, C.F., Verduyn, W., Houwing-Duistermaat, J. *et al.* (2007) A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med.*, **4**, e278.
10. Plenge, R.M., Seielstad, M., Padyukov, L., Lee, A.T., Remmers, E.F., Ding, B., Liew, A., Khalili, H., Chandrasekaran, A., Davies, L.R. *et al.* (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. *N. Engl. J. Med.*, **357**, 1199–1209.
11. Remmers, E.F., Plenge, R.M., Lee, A.T., Graham, R.R., Hom, G., Behrens, T.W., de Bakker, P.I., Le, J.M., Lee, H.S., Batliwalla, F. *et al.* (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.*, **357**, 977–986.
12. Lee, H.S., Remmers, E.F., Le, J.M., Kastner, D.L., Bae, S.C. and Gregersen, P.K. (2007) Association of STAT4 with rheumatoid arthritis in the Korean population. *Mol. Med.*, **13**, 455–460.
13. van der Helm-van Mil, A.H., Huizinga, T.W., de Vries, R.R. and Toes, R.E. (2007) Emerging patterns of risk factor make-up enable subclassification of rheumatoid arthritis. *Arthritis Rheum.*, **56**, 1728–1735.
14. Begovich, A.B., Carlton, V.E., Honigberg, L.A., Schrodi, S.J., Chokkalingam, A.P., Alexander, H.C., Ardlie, K.G., Huang, Q., Smith, A.M., Spoerke, J.M. *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.
15. Wesoly, J., Hu, X., Thabet, M.M., Chang, M., Uh, H., Allaart, C.F., Toes, R.E., Houwing-Duistermaat, J.J., Begovich, A.B. and Huizinga, T.W. (2007) The 620W allele is the PTPN22 genetic variant conferring susceptibility to RA in a Dutch population. *Rheumatology (Oxford)*, **46**, 617–621.
16. Vang, T., Congia, M., Macis, M.D., Musumeci, L., Orru, V., Zavattari, P., Nika, K., Tautz, L., Tasken, K., Cucca, F. *et al.* (2005) Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat. Genet.*, **37**, 1317–1319.
17. Klareskog, L., Stolt, P., Lundberg, K., Kallberg, H., Bengtsson, C., Grunewald, J., Ronnelid, J., Harris, H.E., Ulfgren, A.K., Rantapaa-Dahlqvist, S. *et al.* (2006) 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.
18. Hill, J.A., Southwood, S., Sette, A., Jevnikar, A.M., Bell, D.A. and Cairns, E. (2003) Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. *J. Immunol.*, **171**, 538–541.
19. Becskei, A. and Grusby, M.J. (2007) Contribution of IL-12R mediated feedback loop to Th1 cell differentiation. *FEBS Lett.*, **581**, 5199–5206.
20. Song, H.Y., Rothe, M. and Goeddel, D.V. (1996) The tumor necrosis factor-inducible zinc finger protein A20 interacts with TRAF1/TRAF2 and inhibits NF-kappaB activation. *Proc. Natl Acad. Sci. USA*, **93**, 6721–6725.
21. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.