Open

The Pharmacogenomics Journal (2013) 13, 235 – 241

www.nature.com/tpj

ORIGINAL ARTICLE

Genome-wide association analysis implicates the involvement of eight loci with response to tocilizumab for the treatment of rheumatoid arthritis

J Wang[1,8], AT Bansal[2,8], M Martin[3,8], S Germer[3], R Benayed[3], L Essioux[4], JS Lee[3,9], A Begovich[5], A Hemmings[3], A Kenwright[1], KE Taylor[6], R Upmanyu[1], P Cutler[4], O Harari[1], J Marchini[7], LA Criswell[6] and Adam Platt[1,8,9]

Rheumatoid arthritis (RA) is an immune-mediated inflammatory disease affecting the joints. A heterogeneous response to available therapies demonstrates the need to identify those patients likely to benefit from a particular therapy. Our objective was to identify genetic factors associated with response to tocilizumab, a humanized monoclonal antibody targeting the interleukin (IL)-6 receptor, recently approved for treating RA. We report the first genome-wide association study on the response to tocilizumab in 1683 subjects with RA from six clinical studies. Putative associations were identified with eight loci, previously unrecognized as linked to the IL-6 pathway or associated with RA risk. This study suggests that it is unlikely that a major genetic determinant of response exists, and it illustrates the complexity of performing genome-wide association scans in clinical trials.

The Pharmacogenomics Journal (2013) **13,** 235–241; doi:10.1038/tpj.2012.8; published online 10 April 2012

Keywords: genetic; genome; rheumatoid arthritis; tocilizumab

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic autoimmune disease occurring in approximately 1% of adults.[1] It is an inflammatory synovitis that can lead to joint destruction and physical disability. The presence of autoantibodies (rheumatoid factor (RF) or anticyclic citrullinated peptide), termed seropositivity, occurs in three-quarters of patients.[2] RA has a complex genetic etiology,[3] the strongest known genetic risk factor being HLA-DRB1, and more specifically a group of alleles referred to as the shared epitope,[4] while a further 31 risk loci have also been confirmed in seropositive disease.[5]

Disease-modifying antirheumatic drugs (DMARDs) such as methotrexate are used to treat RA. Second-line therapies include biological agents such as an anti-tumor necrosis factor (anti-TNF) agent, the B-cell depleting agent rituximab, or tocilizumab, a humanized monoclonal antibody that targets the IL-6 receptor. Criteria from the American College of Rheumatology (ACR) define 20, 50 and 70% improvements in disease activity (ACR20, 50 and 70). Response rates to biological therapies are typically 50 – 60% for ACR20, 20 – 40% for ACR50 and 10 – 20% for ACR70.[6] There is considerable heterogeneity of response to a given therapy. Genetic biomarkers might identify which patients are likely to respond to a particular therapy.

Genetic biomarker discovery for response to anti-TNF therapy and rituximab has focused on candidate-based genetic studies[6] of known or hypothetical susceptibility loci or target-related genes. These studies have not yielded a validated biomarker predictive of response to any of these therapies. Using a genome-wide association study (GWAS) approach, we sought to identify genetic polymorphisms that might predict response to tocilizumab and replicate our findings in an independent cohort.

MATERIALS AND METHODS

Subjects

Individuals ($N = 1683$) providing DNA used in this study took part in one of five pivotal phase III studies: RADIATE[7] ($n = 178$); OPTION[8] ($n = 273$); TOWARD[9] ($n = 459$); AMBITION[10] ($n = 247$); LITHE[11] ($n = 469$) or the translational study MEASURE[12] ($n = 57$) (Table 1). These trials evaluated the efficacy and safety of tocilizumab (4 or 8 mg kg^{-1}), administered every 4 weeks over at least a 24-week period compared with DMARD therapy. Methotrexate was the most common DMARD in these trials. The study populations differed according to background therapy with methotrexate or DMARDs (OPTION,[8] TOWARD,[9] LITHE,[11] MEASURE[12]), previous inadequate response to anti-TNF agents (RADIATE[7]), tocilizumab monotherapy in patients with no background DMARD therapy or history of methotrexate inadequate response (AMBITION[10]). DNA samples were collected only from patients who gave separate informed consent to analyses designed to facilitate the study of genetic contributions to specific efficacy and safety responses to tocilizumab. For the discovery of novel variants in the genic regions of SPTLC3 (serine palmitoyltransferase, long-chain base subunit 3) and MYO18B, 194 patients (98 European League Against Rheumatism good responders and 96 European League Against Rheumatism non-responders), matched for age and sex, were selected from Caucasian RA patients treated with tocilizumab (subset 1).

[1]Roche Products Ltd, Welwyn Garden City, UK; [2]Acclarogen Ltd, Cambridge, UK; [3]Roche, Nutley, NJ, USA; [4]Roche, Basel, Switzerland; [5]Roche Diagnostics, Pleasanton, CA, USA; [6]Department of Medicine, Rosalind Russell Medical Research Center for Arthritis, University of California San Francisco, San Francisco, CA, USA and [7]Department of Statistics, University of Oxford, Oxford, UK. Correspondence: Dr J Wang, Roche Products Limited, 6 Falcon Way, Shire Park, Welwyn Garden City AL7 1TW, UK.
E-mail: jianmei.wang@roche.com
[8]These authors contributed equally to this work.
[9]Current address: JS Lee, Montclair, NJ, USA and A Platt, Cheshire, UK.
Received 31 August 2011; revised 23 January 2012; accepted 5 March 2012; published online 10 April 2012

Table 1. Demographic and baseline characteristics of patients

Variable	*OPTION*	*AMBITION*	*RADIATE*	*TOWARD*	*LITHE*	*MEASURE*	*All trials*
Patient population	MTX IR	MTX naïve or free	Anti-TNF IR	DMARD-IR	MTX-IR	DMARD-IR/anti-TNF-IR	
Treatment	8 TCZ+MTX 4 TCZ+MTX MTX	8 TCZ MTX	8 TCZ+MTX 4 TCZ+MTX MTX	8 TCZ+MTX MTX	8 TCZ+MTX 4 TCZ+MTX MTX	8 TCZ+MTX MTX	
Total (*n*)	273	247	178	459	469	57	1683
TCZ+MTX (*n*)	180	108 (mono)	119	299	313	26	1044
MTX (*n*)	93	139	59	160	156	32	669
Race (% White)	76.9	85.4	93.3	77.1	74.2	96.5	79.9
White TCZ+MTX (*n*)	138	85	104	221	215	19	774
White MTX (*n*)	67	113	54	122	112	25	483
Age, mean (s.e.)	51.78 (0.66)	50.14 (0.83)	51.50 (0.93)	54.3 (0.58)	51.84 (0.56)	55.35 (1.31)	52.33 (0.30)
Sex, % female	78.0	76.5	79.8	80.6	83.2	68.4	79.8
Height (cm), mean (s.e.)	163.9 (0.56)	165.7 (0.60)	165.4 (0.62)	163.7 (0.41)	162.3 (0.41)	165.1 (1.35)	163.9 (0.22)
Weight (kg), mean (s.e.)	73.70 (1.06)	75.42 (1.11)	76.72 (1.44)	76.71 (0.84)	73.45 (0.87)	82.08 (2.56)	75.31 (0.45)
RA duration (years), mean (s.e.)	7.75 (0.44)	6.76 (0.51)	11.06 (0.62)	9.79 (0.43)	9.45 (0.37)	8.66 (1.04)	9.02 (0.20)
RF, % positive	75.1	66.4	79.8	75.4	81.5	NA	76.2
Baseline DAS28, mean (s.e.)	6.83 (0.05)	6.77 (0.06)	6.84 (0.07)	6.65 (0.05)	6.59 (0.04)	6.41 (0.12)	6.69 (0.02)
Baseline HAQ mean (s.e.)	1.60 (0.04)	1.51 (0.04)	1.66 (0.04)	1.45 (0.03)	1.52 (0.03)	1.60 (0.08)	1.53 (0.02)
Baseline SJC, mean (s.e.)	20.69 (0.67)	19.70 (0.67)	18.85 (0.80)	20.48 (0.55)	17.37 (0.47)	16.65 (1.30)	19.23 (0.27)
Baseline TJC, mean (s.e.)	32.33 (0.92)	32.60 (0.93)	31.21 (1.14)	30.17 (0.73)	29.04 (0.68)	30.54 (2.19)	30.68 (0.37)
Shared epitope,[a] % positive	71.0	64.6	75.7	72.6	NA	NA	71.0

Abbreviations: DAS, Disease Activity Score; DMARD, disease-modifying antirheumatic drug; HAQ, health assessment questionnaire score; IR, inadequate responder; MTX, methotrexate; NA, not available; RA, rheumatoid arthritis; RF, rheumatoid factor; SJC, swollen joint count; TCZ, tocilizumab; TJC, tender joint count; TNF, tumor necrosis factor.
[a]A subject is positive if he/she carries one or two shared epitope alleles.
8 TCZ, 8 mg kg^{-1} TCZ; 4 TCZ, 4 mg kg^{-1} TCZ.

Genotyping, DNA sequencing and quality control

DNA samples were genotyped using the Illumina Bead-Chip arrays (Illumina, San Diego, CA, USA) as described previously.[13] RADIATE,[7] OPTION[8] and TOWARD[9] were genotyped with the HumanHap550K version 3.0, AMBITION[10] with Human1M-Duo version 3.0, LITHE[11] and MEASURE[12] with HumanOmni1-Quad and sequencing-derived markers with the Illumina custom GoldenGate genotyping assay. Sample quality control (QC) was conducted before genotyping, and data QC was conducted after genotyping (Supplementary Material). Assays with either $>5\%$ missing data or with a minor allele frequency $<1\%$ were excluded from the analysis. χ^2 Tests of Hardy–Weinberg equilibrium were conducted in white subjects; the results were used, along with estimates of minor allele frequency, to assist in the interpretation of associations.

A sequencing approach was used to discover novel variants in two loci of particular interest identified by the *Initial association analysis*, SPTLC3 and MYO18B (Supplementary Material). Genotyping was performed for 1100 variants in SPTLC3 and 1333 variants in MYO18B.

Statistical analysis, principal component analysis and genotype imputation

Principal component analysis was conducted to enable correction for population stratification in the GWAS, using Eigensoft v.3.0.[14,15] Details are provided in Supplementary Material.

Phenotypic outliers were assessed iteratively by calculating the Bonferroni-adjusted *P*-value for the largest absolute studentized deleted residual, using the 'car' package in R v.2.7.2;[16–18] outlying subjects were removed from analyses as appropriate. Single-point association tests were applied in white subjects and separately in subjects of all ethnicities using PLINK v.1.06.[19] Linear or logistic regression modeling of the clinical end point was applied, assuming additive effects for the single-nucleotide polymorphisms (SNPs), and including other co-variates (baseline end point value, dose, ethnicity and study). In association analysis of subjects of all ethnicities, the ethnicity co-variate was represented by the first five principal components from principal component analysis. In association analysis of white subjects, the ethnicity co-variate was represented by indicators of geographic region (Western Europe; North America; South America; Rest of World). In the first hypothesis generating round of analysis, only markers from HumanHap550K v.3.0 were considered. A small number of missing genotypes were imputed using Mendel v.9.0.0,[20,21] without external reference data, in preparation for LASSO (least absolute shrinkage and selection operator) penalized regression.[22] Haplotype frequencies were estimated along a sliding genomic window using a penalized likelihood approach, and the most probable SNP genotype was assigned, based on these estimates. It is noted that if the imputation step had not been included, missing autosomal genotypes would have been replaced by heterozygote calls in Mendel LASSO analysis. LASSO was applied because our univariate analysis might lack power for the discovery of polymorphisms involved in epistatic effects. As a variable selection method, which may be used when the dimensions of the data are greater than the sample size, LASSO penalized regression has been previously applied to several GWAS.[23,24] The penalty factor was adjusted so that 10 SNPs were selected in each analysis.

In the second, confirmatory round of analysis, some SNPs to be confirmed had missing genotypes owing to the different assay platform used. Missing genotypes were imputed using IMPUTE[25,26] before single-point analysis; LASSO analysis was not conducted. In the pooled exploratory analysis, ancestry analysis was repeated as described above for all available subjects and separately for white subjects. In both cases, ProbABEL[27] was used for co-variate-adjusted association testing, to account for uncertainty in imputation.

Study design

Studies were genotyped and analyzed in the order they became available. First, RADIATE,[7] OPTION,[8] TOWARD[9] and AMBITION[10] (ROTA) were used to identify initial associations with six efficacy end points. We chose a continuous clinical end point, change in Disease Activity Score (cDAS28), as a primary end point, as it measures treatment response accounting for the baseline disease activity, and is more sensitive to small effects compared with dichotomous end points (the addition of the prefix 'c' to an abbreviation for a clinical end point denotes 'change in'). As DAS28 is a complex end point, it was considered plausible that a genetic contribution to DAS28 may be more readily detected through association with a particular component of DAS28 or other important measures of disease impact, and thus the following end points were also examined: changes in

Table 2. Numbers of markers identified as significantly associated with efficacy end points

(A) Overlap indicates the number of markers identified in more than one analysis. A marker is selected in White if $P < 10^{-5}$ *in the White population. A marker is selected in All if* $P < 10^{-4}$ *in the White population and a lower* P*-value in the All population*

End point	*White*	*All*	*Overlap, White/All*	*White/All Overlap with LASSO of White*	*White/all Overlap with LASSO of All*	*Overlap LASSO White/LASSO All*	*Total*
cDAS28	5	18	2	3	3	2	33
cSJC	4	25	1	0	2	2	44
cTJC	9	11	1	1	1[a]	1	36
cHAQ	9	13	0	2	2	1	37
cCRP	14	15	0	0	0	2	45
ACR20	3	8	0	5	3	4	21

(B) SNP markers meeting genome-wide significance ($P < 10^{-7}$*)*

End point	*Chr*	*SNP*	*Position*	*White, MAF*	*White HWE* P	*White* β	*White* P	*All* P	*Locus*	*Nearby genes (kb)*
cSJC	2	rs11886534	3 815 178	0.44	0.54	−2.79	5.4e-08	1.4e-07	—	*ALLC*(−87)
cSJC	14	rs850246	56 430 538	0.30	0.63	2.99	7.9e-08	2.6e-06	—	*OTX2*(−88)
cCRP	9	rs13302591	132 054 458	0.16	0.53	0.57	8.2e-08	9.8e-06	—	*FREQ*(−25)
cCRP	6	rs12110787	161 394 956	0.10	0.58	0.54	3.0e-05	3.0e-08	MAP3K4	*AGPAT4*(82)

(C) Associations of SPTLC3 and MYOB18 sequencing-derived markers with cDAS28

Locus	*Chr*	*SNP*	*Position*	*White MAF*	*White HWE* P	*White* P	*All* P
SPTLC3	20	rs6105044	13 065 746	0.32	0.0047	2.3e-05	1.2e-06
SPTLC3	20	rs6041897	13 067 979	0.47	0.59	1.7e-05	3.5e-06
SPTLC3	20	rs6033625	13 073 273	0.32	0.13	1.3e-05	2.1e-06
SPTLC3	20	seq_rs67718002	13 085 393	0.13	0.45	1.3e-05	1.4e-06
MYOB18	22	rs6004918	24 784 171	0.18	0.20	1.2e-06	0.0007

Abbreviations: ACR, American College of Rheumatology; cCRP, change in C-reactive protein; cDAS28, change in Disease Activity Score; cHAQ, change in health assessment questionnaire score; cSJC, change in swollen joint count; cTJC, change in tender joint count; HWE, Hardy–Weinberg equilibrium; LASSO, least absolute shrinkage and selection operator; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.
[a]Only nine predictors selected in cTJC LASSO of All.
For details on individual markers see Supplementary Table 2.

swollen joint count (cSJC), tender joint count (cTJC), health assessment questionnaire score (cHAQ) and C-reactive protein (cCRP). As these end points, particularly the components of DAS, are not independent, the multiple testing burden is not as great compared to an assessment of multiple, independent outcomes. As week 16 is the last data time point before escape therapy, it was used to maximize subject numbers, with the exception of cCRP, where week 8 was used as a steady state has been reached by this time point. ACR20 at week 24 was also examined as it was the primary end point in all studies. Candidate SNPs were selected as meeting at least one of the following criteria: (1) $P < 10^{-5}$ in white subjects (White) as they were the largest ethnic group with recognized low ethnic confounding; (2) $P < 10^{-4}$ in White and a lower *P*-value in all ethnicities (All); or (3) selected by LASSO analysis of White or All. These non-conservative thresholds are above a genome-wide significance level of $P \leqslant 10^{-7}$ and reflect a greater emphasis on power than type I error for the purpose of hypotheses generation (Supplementary Material).

Two loci of particular interest from this initial association analysis were also sequenced in a subset of patients; novel variants were then genotyped in ROTA.[7–10] Association candidates were selected using criteria 1 and 2 above.

Association candidates from analysis of ROTA[7–10] underwent confirmatory genotyping in an independent cohort consisting of LITHE[11] and MEASURE[12] (LM). For those candidates not directly genotyped owing to the change in assay platform, imputed genotype probabilities were used. If imputed data were not available, the proxy SNP with the highest linkage disequilibrium r^2 that passed QC was utilized. Markers in either White or All subjects with $P < 0.05$ against the same end point as the original association or $P < 0.05$ against the primary end point cDAS in both LM[11,12] and ROTA,[7–10] with the same directionality, were considered to have achieved confirmation. Estimations of effect sizes are presented for DMARD inadequate responder (IR) patients represented from LITHE,[11] OPTION[8] and TOWARD[9] (LOT), anti-TNF IR patients from RADIATE[7] and methotrexate-naïve/free patients from AMBITION.[10] Replicated SNPs were also examined for predictive effect size on DAS28 remission (DAS28 <2.6) in singular, pair-wise and triple combinations in both additive and multiplicative models. Predictive vs prognostic property was investigated via comparison of effect on tocilizumab- and placebo-treated patients.

To utilize the greater sample size afforded through combination of all studies, an additional analysis explored the association of cDAS28, cTJC, cSJC and cHAQ with the union of imputed genotypes from ROTA[7–10] and LITHE[11] (ROTAL). To investigate the influence of RF on the predictive or prognostic nature of candidate SNPs, MEASURE[12] was excluded from this analysis as RF data were not collected in this study.

RESULTS

Initial association analysis

In ROTA,[7–10] 42 patients failed genotyping QC, leaving 1157 individuals, of whom 706 were treated with tocilizumab. Of the 534 053 markers analyzed, 207 markers were identified as having 253 significant associations in tocilizumab-treated patients as

defined by our selection criteria (Supplementary Table 2). Of these 207 markers, 88 were identified only through LASSO analysis. Four SNPs that met a genome-wide significance threshold of $P \leqslant 10^{-7}$ and a summary of univariate associations generated per end point are provided in Table 2. HLA-DRB1 (shared epitope) was also genotyped in ROTA,[7–10] but no association with tocilizumab response was found (Supplementary Table 3).

Two SNP markers were highlighted as of particular interest. rs6078937, an intronic SNP in SPTLC3, was the only marker highlighted by all four analysis criteria for any end point (cDAS: $P = 4.43 \times 10^{-6}$ in White, $P = 1.73 \times 10^{-6}$ in All and selected by LASSO in both White and All). Another SNP, rs6004913, was the only marker highlighted for four different end points: in the White population, cDAS28: $P = 3.94 \times 10^{-6}$, cTJC: $P = 9.71 \times 10^{-6}$ (plus 3 flanking SNPs) and cHAQ: $P = 1.50 \times 10^{-6}$ (plus 1 flanking SNP); and in the All population, ACR20: $P = 4.36 \times 10^{-5}$ (plus 4 flanking SNPs). The variant rs6004913 is in linkage disequilibrium with rs2236006 with $r^2 = 0.62$, a non-synonymous coding change in MYO18B (MYOSIN XVIIIB). To identify potential causal variants, we re-sequenced these loci in a 194 subject responder/non-responder subset (subset 1) revealing 1100 and 1333 unique variants in SPTLC3 and MYO18B, respectively. These variants were then genotyped in ROTA,[7–10] with 241 SPTLC3 and 167 MYO18B markers having a minor allele frequency $\geqslant 0.05$. Of these 408 variants, five preliminary associations were identified (Table 2C).

Confirmatory analysis

In all, 253 initial associations together with five sequencing variant associations were tested in an independent group of subjects, LM.[11,12] Four LM[11,12] patients failed genotyping QC, leaving 526 individuals, of whom 338 were treated with tocilizumab. In total, 371 subjects were defined as White as described in Figure 1, of whom 234 were treated with tocilizumab. Of 208 SNPs, 127 were directly genotyped, while 79 SNPs were imputed and two proxies were used. The White tocilizumab-treated population yields 80% power to confirm preliminary ROTA[7–10] associations, if the variant has a minor allele frequency of at least 0.2, expresses an additive effect of at least 0.35 units of change in DAS28, using the significance threshold of 0.05 (Supplementary Table 1).

Seven markers, five directly genotyped and two imputed, achieved confirmation at $P < 0.05$ for the same end point with which they were identified in ROTA[7–10] (Table 3). Three of these were also confirmed against cDAS28, while a fourth marker, rs7055107, represented by a proxy, was confirmed against cDAS28 in LM.[11,12] Interestingly, all of these eight markers originated from LASSO regression; no markers generated from sequencing were confirmed. The observed association with efficacy could potentially be characterized as predictive, that is, associated specifically with response to tocilizumab, and not to placebo plus DMARD background therapy. Alternatively, the associations could be prognostic, that is, associated with a change in disease activity through a spontaneous disease fluctuation or general treatment effect, or may be dependent on dose of tocilizumab, or specific to treatment line. To address these possibilities, we analyzed the predictive nature of these polymorphisms on cDAS28 in various genetic models (Figure 2 and Supplementary Figure 1) and found them to be largely predictive. As a major aim in the treatment of RA is to attain remission, we also analyzed the predictive nature of these polymorphisms in terms of DAS28 < 2.6 both singularly and in combination (Table 4).[28]

Pooled exploratory analysis

ROTAL[7–11] consisted of 1626 individuals passing genotyping QC (All). Of 1091 patients treated with tocilizumab, 791 were RF positive. Principal component analysis was used to identify a

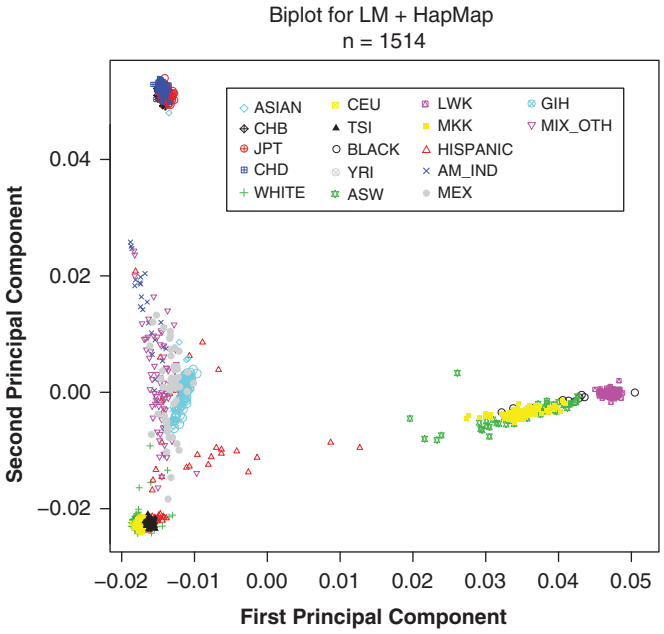

Figure 1. Biplot of the first two principal components, utilizing 526 LITHE and MEASURE (LM)[11,12] subjects and 988 founder subjects. Five LM[11,12] subjects, self-reporting as White and non-Hispanic and represented by green crosses, fell outside the main cluster of Caucasian subjects. Furthermore, 59 subjects, self-reporting as Hispanic and represented by red triangles, fell within the main Caucasian cluster. For the purposes of analysis, a threshold of −0.02 was placed on the second principal component; subjects falling below the threshold were considered to form a homogeneous group and were incorporated into White.

753-subject tocilizumab-treated Caucasian subgroup (White), of whom 573 were RF positive. Principal components 1–5, as well as dose, study and baseline measure of the corresponding end point, were included in the corresponding regression model as co-variates. A total of 1 470 186 markers were tested for single-point association with response in RF-positive and -negative groups for both All and White populations, 42 markers passing significance levels of $P < 10^{-6}$ with a frequency $> 1\%$ (Supplementary Table 5). Three markers met genome-wide significance at $P < 10^{-7}$ and frequency $> 1\%$; all were in the White RF-negative group (Supplementary Table 6A). However, as they were identified in the smallest subgroup ($n = 179$), the numbers of patients carrying the minor alleles ($n = 46$, 24 and 6, respectively) were considered too small to be of further interest at this stage. Interestingly, 13 markers in linkage disequilibrium with ENOX1, a locus highlighted by rs9594987 confirmation in LM[11,12] (Table 3), were associated with cHAQ in the RF-positive All population, seven at $P < 10^{-4}$ and six at $P < 10^{-6}$ (Supplementary Table 6B). Associations were weaker in All RF-negative ($P > 0.05$), White RF-positive ($P > 10^{-4}$) and White RF-negative ($P > 0.01$) populations.

Of seven markers achieving confirmation in LM[11,12] (Table 3), three demonstrated a differential association to response in ROTAL[7–11] according to seropositivity. (rs7055107 was not considered as X-chromosome markers were not imputed.) rs4910008 is associated more strongly with cTJC, cSJC and cDAS28, rs9594987 with cHAQ and rs703505 with cHAQ, cTJC and cDAS28, all in RF-positive patients (Supplementary Table 6C). RA risk loci[5] were also examined in seropositive white patients (Supplementary Table 4). All observed associations were weak and would not surpass an adjustment for multiple testing, although rs3093023, from the CCR6 locus, did associate at $P < 0.05$ to cDAS, cSJC and cTJC.

Table 3. Markers achieving confirmation in LM[11,12] either (A) to the end point with which they were associated in ROTA[7–10] or (B) to cDAS28, either genotyped directly (direct), genotype imputed (Impute) or proxy SNP utilized (proxy)

(A)

ROTA marker	*Pop*	*MAF*	*End point*	P-*value*	*Chr*	*Position*	*Coding*	*Info*	β *or* **OR*	*TCZ* P
rs11052877	W	0.38	ACR20_WK24	LASSO	12	9 796 957	CD69	Direct	0.56*	0.0039
rs4910008	W	0.47	cTJC_WK16	LASSO	11	11 436 442	GALNTL4	Direct	−3.28	0.0063
rs9594987	A	0.44	cHAQ_WK16	LASSO	13	43 128 994	ENOX1	Impute (0.996)	−0.10	0.016
rs10108210	A	0.41	cHAQ_WK16	LASSO	8	2 588 854	—	Impute (0.973)	0.09	0.028
rs703297	A	0.48	ACR20_WK24	LASSO	2	19 489 299	—	Direct	0.68*	0.022
rs703505	A	0.42	cHAQ_WK16	LASSO	5	169 741 980	KCNIP1	Direct	−0.09	0.031
rs1560011	A	0.42	ACR20_WK24	LASSO	12	9 714 219	CLEC2D	Direct	0.72*	0.046

(B)

Marker	*Pop*	*MAF*	*End point*	*ROTA, assoc.*	*Chr*	*Position*	*Coding/nearby (kb)*	*Info*	*CDAS28 week 16*			
									ROTA TCZ		*LM TCZ*	
									P-*value*	β	P-*value*	β
rs703505	A	0.42	cHAQ_WK16	LASSO	5	169 741 980	KCNIP1	Direct	0.001	−0.26	0.0002	−0.45
rs4910008	W	0.47	cTJC_WK16	LASSO	11	11 436 442	GALNTL4	Direct	0.01	−0.24	0.04	−0.28
rs11052877	W	0.38	ACR20_WK24	LASSO	12	9 796 957	CD69	Direct	0.01	0.21	0.05	0.23
rs7055107	A	0.48	ACR20_WK24	LASSO	23	46 456 428	SLC9A7	Proxy	0.006	−0.28	0.05	−0.21

Abbreviations: A, All; cDAS28, change in Disease Activity Score; LASSO, least absolute shrinkage and selection operator; LM, LITHE and MEASURE; MAF, minor allele frequency; OR, odds ratio; Pop, population; ROTA, RADIATE, OPTION, TOWARD and AMBITION; SNP, single-nucleotide polymorphism; TCZ, tocilizumab; W, White.
Imputation scores are given within parentheses.

DISCUSSION

We conducted the first genetic analysis of response to tocilizumab, revealing putative associations with eight loci (Table 3, Figure 2 and Supplementary Figure 1), none of which have been previously implicated as risk alleles for RA or associated with response to any other therapy.[5] None of these are obviously linked to the IL-6 pathway, and there is no association of shared epitope with tocilizumab response. Several of these associations were stronger in seropositive patients (Supplementary Table 6C), in part because the majority of patients are seropositive.

There is a clear difference in DAS28 remission rates between candidate genetic biomarker-positive and -negative patients (Table 4). However, the differences are modest, capturing an insufficient portion (<2%) of heterogeneity in response to be clinically useful.

These polymorphisms may help to reveal insights into the mechanism of action of tocilizumab in RA. Of the polymorphisms confirmed in the LM data set,[11,12] SNPs from two genes encoding C-lectins, that is, part of the NK gene complex on chromosome 12,[29] were highlighted. rs1560011 is an intronic SNP in CLEC2D, and rs11052877 is located in the 3′-untranslated region of CD69 (CLEC2C). Figure 2 demonstrates tocilizumab-treated patients carrying the major allele of either polymorphism have a better DAS28 response in a recessive model; this is not observed with placebo. Without functional analysis we can only hypothesize how variation within CLEC2D and CD69 might influence response. The receptor, expressed on NK cells, T cells, activated B cells and dendritic cells, induces NK cells to produce interferon-γ.[30] CLEC2D blocks osteoclast formation.[31,32] The ligand of CLEC2D, KLRB1 (CD161), identifies IL-17-producing T-cell subsets.[33] CD69 is expressed on a number of hematopoietic cells, and crosslinking CD69 induces proliferation, IL-2 and interferon-γ release and NO production.[34]

The biological significance of genes represented by the other markers is not obvious, but these polymorphisms identified may provide clues to the molecular basis for tocilizumab response in RA. With a pooled exploratory analysis utilizing the ROTAL data set,[7–11] we have identified a further number of polymorphisms putatively associated with response. Three of these polymorphisms met genome-wide significance thresholds but were present in a small number of patients and so will require further validation.

Only analyses of efficacy end points were reported in this manuscript. Using a candidate gene approach, bilirubin elevation was previously shown to be strongly associated with genetic variations within UGT1A1.[13] For other safety end points, such as neutropenia, genetic analysis was not undertaken owing to the very low event rate.

Our analyses illustrate the complexity of performing genome-wide association scan in clinical trials. Despite having over 1600 patients from six randomized, controlled clinical studies and rigorous methodology, only putative loci of weak associations with efficacy were identified. This indicates that treatment response in RA is likely to be complex, even for a targeted biologic. It is possible that association might be stronger in more homogeneous subgroups of patients (such as methotrexate-naïve RA treated with tocilizumab as monotherapy[10]), or by comparing only the extremely good with the extremely poor responders. However, our GWAS has limited utility in these contexts owing to small subgroup sizes (see Supplementary Material).

There are several limitations with our study. A two-stage analysis was performed. In the SNP discovery stage, we lowered the threshold below genome-wide significance and analyzed multiple efficacy outcome measures: DAS28 and core components of disease activity and response (SJC, TJC, HAQ and CRP).[35] We combined selection criteria based on univariate and multivariate approaches, to increase our chances of identifying polymorphisms

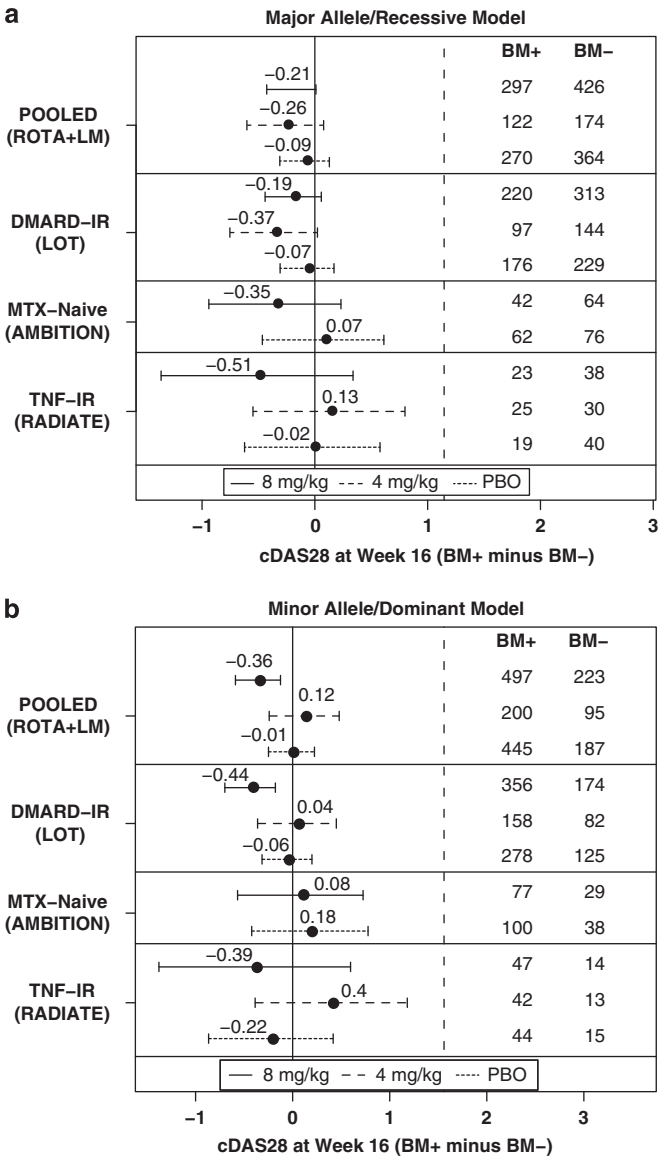

Figure 2. Effect of single-nucleotide polymorphism (SNP) variants (**a**) rs11052877 (CD69), (**b**) rs4910008 (GALNTL4) on change in Disease Activity Score (cDAS28) in response to tocilizumab in All. Biomarker-positive/negative patients are defined as those carrying the major or minor allele and genetic model as indicated in each forest plot. The 95% confidence intervals are indicated for the point estimate of the difference in mean cDAS28 between biomarker-positive and -negative patients. Values are presented for all subjects and separated by treatment line, dose and drug. Interpretation of effect size estimations for patients previously exposed to anti-tumor necrosis factor (anti-TNF) inadequate responder (IR) (RADIATE[7]) or methotrexate naïve (AMBITION[10]) are confounded by the low number of patients in this cohort leading to broad confidence intervals, but are presented for completeness. BM+, biomarker positive, BM−, biomarker negative.

that contribute to any epistatic effect on overall DAS28 response. Such a multi-pronged strategy increases type I error, but interestingly, the majority of SNPs confirmed in LM[11,12] were revealed by LASSO analysis. In the confirmatory stage, the sample size of the cohort (LM) was relatively small, thus limiting power (Supplementary Table 1), and a non-conservative threshold ($P = 0.05$) was used. Our aim to identify predictive biomarkers specific for tocilizumab response was challenged by the lack of

Table 4. DAS28 remission rates at week 16 for DMARD IR (LOT) patients stratified by SNPs achieving confirmation in LM

Marker	*LOT freq*	*DAS28 remission rate (%)*		
		BM+	*BM−*	*Δ*
rs7055107	0.30	28.4	18.9	9.4
rs7055107 or rs9594987	0.45	28.4	16.2	12.2
rs7055107 and rs4910008	0.21	33.3	18.8	14.5
rs7055107 or rs9594987 or rs11052877	0.78	24.4	12.5	11.9
rs7055107 and rs703505 and rs10108210	0.20	31.8	19.2	12.6

Abbreviations: BM+, biomarker positive; BM−, biomarker negative; DAS28, Disease Activity Score; DMARD, disease-modifying antirheumatic drug; IR, inadequate responder; LM, LITHE and MEASURE; LOT, LITHE, OPTION, and TOWARD; SNP, single-nucleotide polymorphism.
Placebo-corrected remission rates and frequency are shown from patients carrying polymorphisms (biomarker positive) and those not carrying the polymorphisms (biomarker negative). Delta (Δ) represents the difference in remission rate between BM+ and BM−. Only combinations yielding the best Δ for each model are shown. Only 8 mg kg^{-1} tocilizumab-treated patients were considered in this analysis owing to the far higher remission rates observed with this dose.[28]

an active comparator drug within the available studies (the comparator arms were generally on matched background therapy with a placebo). Finally, there was no independent cohort in which to confirm the combinations of SNPs selected to predict DAS28 remission rates. Taken together, these limitations suggest that the eight reported associations require confirmation in an independent cohort before they can be considered validated. Our findings demonstrate the need for very large, adequately powered cohorts to conduct robust pharmacogenetic response analyses in a heterogeneous autoimmune disease such as RA.

CONFLICTS OF INTEREST

JW, MM, SG, RB, LE, JL, AB, AH, AK, RU, PC, OH and AP are employees of Roche. ATB and JM received compensation as consultants for Roche.

ACKNOWLEDGEMENTS

We thank Teodorica Bugawan of Roche Molecular Diagnostics for providing the shared epitope data, and Peter K Gregersen of The Feinstein Institute for Medical Research for his advice and contributions during the early design phase of the GWAS. The registration names and numbers of tocilizumab clinical trials on Clinicaltrials.gov are: OPTION, NCT00106548; TOWARD, NCT00106574; RADIATE, NCT00106522; AMBITION, NCT00109408; LITHE, NCT00106535; and MEASURE, NCT00535782.

REFERENCES

1 Gabriel SE, Crowson CS, O'Fallon WM. The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955 – 1985. *Arthritis Rheum* 1999; **42**: 415 – 420.
2 Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. *Lancet* 2009; **373**: 659 – 672.
3 Newton JL, Harney SM, Wordsworth BP, Brown MA. A review of the MHC genetics of rheumatoid arthritis. *Genes Immun* 2004; **5**: 151 – 157.
4 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 – 1213.
5 Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010; **42**: 508 – 514.
6 Verweij CL. Pharmacogenetics: anti-TNF therapy in RA—towards personalized medicine? *Nat Rev Rheumatol* 2011; **7**: 136 – 138.
7 Emery P, Keystone E, Tony HP, Cantagrel A, van Vollenhoven R, Sanchez A *et al.* IL-6 receptor inhibition with tocilizumab improves treatment outcomes in patients with rheumatoid arthritis refractory to anti-tumour necrosis factor

biologicals: results from a 24-week multicentre randomised placebo-controlled trial. *Ann Rheum Dis* 2008; **67**: 1516 – 1523.
8 Smolen JS, Beaulieu A, Rubbert-Roth A, Ramos-Remus C, Rovensky J, Alecock E *et al.* Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (OPTION study): a double-blind, placebo-controlled, randomised trial. *Lancet* 2008; **371**: 987 – 997.
9 Genovese MC, McKay JD, Nasonov EL, Mysler EF, da Silva NA, Alecock E *et al.* Interleukin-6 receptor inhibition with tocilizumab reduces disease activity in rheumatoid arthritis with inadequate response to disease-modifying antirheumatic drugs: the tocilizumab in combination with traditional disease-modifying antirheumatic drug therapy study. *Arthritis Rheum* 2008; **58**: 2968 – 2980.
10 Jones G, Sebba A, Gu J, Lowenstein MB, Calvo A, Gomez-Reino JJ *et al.* Comparison of tocilizumab monotherapy versus methotrexate monotherapy in patients with moderate to severe rheumatoid arthritis: the AMBITION study. *Ann Rheum Dis* 2010; **69**: 88 – 96.
11 Kremer JM, Blanco R, Brzosko M, Burgos-Vargas R, Halland AM, Vernon E *et al.* Tocilizumab inhibits structural joint damage in rheumatoid arthritis patients with inadequate responses to methotrexate: results from the double-blind treatment phase of a randomized placebo-controlled trial of tocilizumab safety and prevention of structural joint damage at one year. *Arthritis Rheum* 2011; **63**: 609 – 621.
12 McInnes IB, Lee JS, Wu W, Giles JT, Bathon JM, Salmon JE *et al.* Lipid and inflammation parameters: a translational, randomized placebo-controlled study to evaluate effects of Tocilizumab: the MEASURE study [abstract]. *Arthritis Rheum* 2010; **62**(Suppl 10): 1441.
13 Lee JS, Wang J, Martin M, Germer S, Kenwright A, Benayed R *et al.* Genetic variation in UGT1A1 typical of Gilbert syndrome is associated with unconjugated hyperbilirubinemia in patients receiving tocilizumab. *Pharmacogenet Genomics* 2011; **21**: 365 – 374.
14 Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006; **2**: e190.
15 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904 – 909.
16 Cook RD, Weisberg S. *Residuals and Influence in Regression: Monographs on Statistics and Applied Probability*. Chapman & Hall: New York, 1982: 1 – 240.
17 Fox J. *Applied Regression Analysis and Generalized Linear Models*, 2nd edn. Sage Publications: Thousand Oaks, CA, 2008.
18 R Development Core Team. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2008.
19 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559 – 575.
20 Lange K, Cantor R, Horvath S, Perola M, Sabatti C, Sinsheimer J *et al.* Mendel version 4.0: a complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. *Am J Hum Genet* 2001; **69**: 504.
21 Wu TT, Lange K. Coordinate descent algorithms for lasso penalized regression. *Ann Appl Stat* 2008; **2**: 224 – 244.
22 Tibshirani R. Regression shrinkage and selection via the lasso. *J R Statist Soc Ser B* 1996; **58**: 267 – 288.
23 Wu TT, Chen YF, Hastie T, Sobel E, Lange K. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* 2009; **25**: 714 – 721.
24 Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am J Hum Genet* 2010; **86**: 6 – 22.
25 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.
26 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906 – 913.
27 Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinform* 2010; **11**: 134.
28 Smolen JS, Aletaha D. Interleukin-6 receptor inhibition with tocilizumab and attainment of disease remission in rheumatoid arthritis: the role of acute-phase reactants. *Arthritis Rheum* 2011; **63**: 43 – 52.
29 Sattler S, Ghadially H, Reiche D, Karas I, Hofer E. Evolutionary development and expression pattern of the myeloid lectin-like receptor gene family encoded within the NK gene complex. *Scand J Immunol* 2010; **72**: 309 – 318.
30 Bambard ND, Mathew SO, Mathew PA. LLT1-mediated activation of IFN-gamma production in human natural killer cells involves ERK signalling pathway. *Scand J Immunol* 2010; **71**: 210 – 219.
31 Hu YS, Zhou H, Myers D, Quinn JM, Atkins GJ, Ly C *et al.* Isolation of a human homolog of osteoclast inhibitory lectin that inhibits the formation and function of osteoclasts. *J Bone Miner Res* 2004; **19**: 89 – 99.
32 Zhou H, Kartsogiannis V, Hu YS, Elliott J, Quinn JM, McKinstry WJ *et al.* A novel osteoblast-derived C-type lectin that inhibits osteoclast formation. *J Biol Chem* 2001; **276**: 14916 – 14923.
33 Maggi L, Santarlasci V, Capone M, Peired A, Frosali F, Crome SQ *et al.* CD161 is a marker of all human IL-17-producing T-cell subsets and is induced by RORC. *Eur J Immunol* 2010; **40**: 2174 – 2181.
34 Marzio R, Mauel J, Betz-Corradin S. CD69 and regulation of the immune function. *Immunopharmacol Immunotoxicol* 1999; **21**: 565 – 582.
35 Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995; **38**: 44 – 48.

Supplementary Information accompanies the paper on the The Pharmacogenomics Journal website (http://www.nature.com/tpj)