nature genetics

Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci

Eli A Stahl[1,2], Soumya Raychaudhuri[1–3], Elaine F Remmers[4], Gang Xie[5], Stephen Eyre[6], Brian P Thomson[2], Yonghong Li[7], Fina A S Kurreeman[1,2,8], Alexandra Zhernakova[9], Anne Hinks[6], Candace Guiducci[2], Robert Chen[1], Lars Alfredsson[10], Christopher I Amos[11], Kristin G Ardlie[2], BIRAC Consortium[33], Anne Barton[6], John Bowes[6], Elisabeth Brouwer[12], Noel P Burtt[2], Joseph J Catanese[7], Jonathan Coblyn[1], Marieke J H Coenen[13], Karen H Costenbader[1], Lindsey A Criswell[14], J Bart A Crusius[15], Jing Cui[1], Paul I W de Bakker[2,16], Philip L De Jager[2,17], Bo Ding[10], Paul Emery[18], Edward Flynn[6], Pille Harrison[19], Lynne J Hocking[20], Tom W J Huizinga[8], Daniel L Kastner[4], Xiayi Ke[6], Annette T Lee[21], Xiangdong Liu[5], Paul Martin[6], Ann W Morgan[18], Leonid Padyukov[22], Marcel D Posthumus[12], Timothy R D J Radstake[23], David M Reid[20], Mark Seielstad[24], Michael F Seldin[25], Nancy A Shadick[1], Sophia Steer[26], Paul P Tak[27], Wendy Thomson[6], Annette H M van der Helm-van Mil[8], Irene E van der Horst-Bruinsma[28], C Ellen van der Schoot[29], Piet L C M van Riel[23], Michael E Weinblatt[1], Anthony G Wilson[30], Gert Jan Wolbink[29,31], B Paul Wordsworth[19], YEAR Consortium[33], Cisca Wijmenga[9], Elizabeth W Karlson[1], Rene E M Toes[8], Niek de Vries[27], Ann B Begovich[7,32], Jane Worthington[6,34], Katherine A Siminovitch[5,34], Peter K Gregersen[21,34], Lars Klareskog[22,34] & Robert M Plenge[1,2]

© 2010 Nature America, Inc. All rights reserved.

To identify new genetic risk factors for rheumatoid arthritis, we conducted a genome-wide association study meta-analysis of 5,539 autoantibody-positive individuals with rheumatoid arthritis (cases) and 20,169 controls of European descent, followed by replication in an independent set of 6,768 rheumatoid arthritis cases and 8,806 controls. Of 34 SNPs selected for replication, 7 new rheumatoid arthritis risk alleles were identified at genome-wide significance ($P < 5 \times 10^{-8}$) in an analysis of all 41,282 samples. The associated SNPs are near genes of known immune function, including *IL6ST*, *SPRED2*, *RBPJ*, *CCR6*, *IRF5* and *PXK*. We also refined associations at two established rheumatoid arthritis risk loci (*IL2RA* and *CCL21*) and confirmed the association at *AFF3*. These new associations bring the total number of confirmed rheumatoid arthritis risk loci to 31 among individuals of European ancestry. An additional 11 SNPs replicated at $P < 0.05$, many of which are validated autoimmune risk alleles, suggesting that most represent genuine rheumatoid arthritis risk alleles.

Rheumatoid arthritis is a common autoimmune disease that affects up to 1% of the general adult population worldwide[1]. Approximately two-thirds of cases are seropositive for rheumatoid factor or anti-cyclic

[1]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA. [2]Broad Institute, Cambridge, Massachusetts, USA. [3]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [4]Genetics and Genomics Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, US National Institutes of Health, Bethesda, Maryland, USA. [5]Department of Medicine, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, Ontario, Canada. [6]Arthritis Research UK Epidemiology Unit, Stopford Building, The University of Manchester, Manchester, UK. [7]Celera, Alameda, California, USA. [8]Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands. [9]Department of Genetics, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands. [10]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [11]University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA. [12]Department of Rheumatology and Clinical Immunology, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands. [13]Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [14]Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California at San Francisco, San Francisco, California, USA. [15]Laboratory of Immunogenetics, Department of Pathology, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. [16]Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts, USA. [17]Department of Neurology, Center for Neurologic Diseases, Brigham and Women's Hospital, Boston, Massachusetts, USA. [18]National Institute for Health Research-Leeds Musculoskeletal Biomedical Research Unit, Leeds Institute of Molecular Medicine, University of Leeds, Leeds, UK. [19]University of Oxford Institute of Musculoskeletal Sciences, Botnar Research Centre, Oxford, UK. [20]Musculoskeletal and Genetics Section, Division of Applied Medicine, University of Aberdeen, Aberdeen, UK. [21]The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, USA. [22]Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden. [23]Department of Rheumatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [24]Genome Institute of Singapore, Singapore. [25]Rowe Program in Genetics, University of California at Davis, Davis, California, USA. [26]Clinical and Academic Rheumatology, Kings College Hospital National Health Service Foundation Trust, Denmark Hill, London, UK. [27]Clinical Immunology and Rheumatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. [28]Department of Rheumatology, Vrije Universiteit University Medical Center, Amsterdam, The Netherlands. [29]Sanquin Research Landsteiner Laboratory, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. [30]School of Medicine and Biomedical Sciences, Sheffield University, Sheffield, UK. [31]Jan van Breemen Institute, Amsterdam, The Netherlands. [32]Roche Diagnostics, Pleasanton, California, USA. [33]A full list of members is provided in the **Supplementary Note**. [34]These authors contributed equally to this work. Correspondence should be addressed to R.M.P. (rplenge@partners.org).

Received 24 November 2009; accepted 25 February 2010; published online 9 May 2010; doi:10.1038/ng.582

Table 1 Study subjects

	Case collection	Control collection	Geographical origin	Case antibody status	Cases	Controls	Genotyping platform	Case-control stratification correction
Meta-analysis[a]	Brigham Rheumatoid Arthritis Sequential Study (BRASS)	Shared controls	Boston, USA	100% CCP+	483	1,449	Affymetrix 6.0	GWAS data PC matching
	CANADA	CANADA and Shared controls	Toronto, Canada	100% CCP+	589	1,472	Illumina 370K	GWAS data PC matching
	Epidemiological Investigation of Rheumatoid Arthritis (EIRA)	EIRA	Sweden	100% CCP+	1,173	1,089	Illumina 317K	Epidemiologically matched & GWAS data PC matching
	North American Rheumatoid Arthritis Consortium (NARAC) I	Shared controls	North America	100% CCP+	867	1,041	Illumina 550K	GWAS data PC matching
	NARAC III	Shared controls	North America	100% CCP+	902	4,510	Illumina 317K	GWAS data PC matching
	Wellcome Trust Case Control Consortium (WTCCC)	Shared controls from WTCCC	United Kingdom	100% RF+ or CCP+	1,525	10,608	Affymetrix 500K	Geographically matched
Replication[b]	CANADA II	CANADA II	Toronto and Halifax, Canada	100% CCP+	1,076	1,269	Sequenom iPlex	Geographically matched
	Dutch	Dutch	The Netherlands	100% RF+	718	697	Sequenom iPlex	Geographically matched
	Genetics Network Rheumatology Amsterdam (GENRA)	GENRA	Amsterdam, The Netherlands	100% CCP+	519	1,155	Sequenom iPlex	Geographically matched
	Genomics Collaborative Initiative (GCI)	GCI	North America	100% RF+	461	460	Kinetic PCR	Epidemiologically matched
	Leiden University Medical Center (LUMC)	LUMC	Leiden, The Netherlands	100% RF+ or CCP+	310	544	Kinetic PCR	Geographically matched
	NARAC II	Shared controls	North America	100% CCP+	462	693	Sequenom iPlex	Ancestry informative marker data matching
	United Kingdom Rheumatoid Arthritis Genetics (UKRAG)	UKRAG	United Kingdom	100% RF+ or CCP+	2,906	3,494	Sequenom iPlex	Geographically matched
	Nurses Health Study (NHS)	NHS	North America	100% RF+ or CCP+	316	494	Biotrove OpenArray	Epidemiologically matched

[a]Meta-analysis; 5,539 cases, 20,169 controls. [b]Replication; 6,768 cases, 8,806 controls. Meta-analysis of GWAS results for six collections (top panel) was used to identify SNPs for replication in eight collections (bottom panel). For each collection, we list the source of controls, geographic origin, autoantibody status of cases, numbers of cases and controls, genotyping platform (GWAS collection microarray platform and replication and validation/prediction collection direct-genotyping technology) and the strategy used to correct for case-control population stratification. See Online Methods and **Supplementary Note** for details. RF, rheumatoid factor; CCP, cyclic citrullinated peptide; PC, principal components.

citrullinated peptide (anti-CCP) autoantibodies[2]. Genetic studies in autoantibody-positive rheumatoid arthritis among subjects of European ancestry have identified multiple risk alleles in the major histocompatibility complex (MHC) region, as well as 25 confirmed rheumatoid arthritis risk alleles in 23 non-MHC loci[3–15]. These alleles explain about 23% of the genetic burden of rheumatoid arthritis[11], indicating that additional alleles remain to be discovered.

To identify new rheumatoid arthritis risk alleles, we conducted a GWAS meta-analysis of 5,539 autoantibody-positive rheumatoid arthritis cases and 20,169 controls of European ancestry (**Table 1**). This study expands upon our previous GWAS meta-analysis of 3,393 cases and 12,462 controls[11] by (i) adding a new GWAS dataset of 483 rheumatoid arthritis cases recruited from the Boston area (Brigham Rheumatoid Arthritis Sequential Study, BRASS) and 1,449 shared controls, (ii) adding 513 cases and 431 controls recruited from Sweden (Epidemiological Investigation of Rheumatoid Arthritis, EIRA), (iii) incorporating a recently published GWAS of 2,418 cases and 4,504 controls recruited from North America (Canada and North American Rheumatoid Arthritis Consortium-III, NARAC-III)[13], and (iv) adding additional shared controls to the NARAC-III dataset. For each of six GWAS case-control collections, we removed SNPs and individuals that failed quality control, matched case-control samples using principal-components analysis to minimize bias due to stratification, and imputed[16] genome-wide to infer genotypes at additional European (CEU) HapMap SNPs. We used logistic regression to conduct a GWAS for 2.56 million SNPs in each collection, corrected for residual inflation using genomic control[17], and combined the results across collections by inverse variance-weighted meta-analysis to calculate P_{GWAS}[18]. We found little evidence of systematic bias as indicated by the genomic control inflation factor ($\lambda_{GC} = 1.04$)[17], the quantile-quantile plots, the results for markers highly differentiated across Europe[19] or comparisons with an alternative analysis using principal-components analysis eigenvectors as covariates (**Supplementary Figs. 1–4** and **Supplementary Table 1**; see Online Methods and **Supplementary Note** for full details of the analysis).

Our GWAS meta-analysis found support for rheumatoid arthritis risk loci previously confirmed among individuals of European ancestry, consistent with power for our study design (**Table 2** and **Supplementary Fig. 5**). Four of the 25 confirmed non-MHC risk alleles obtained $P_{GWAS} < 5 \times 10^{-8}$ (at the *PTPN22*, *CTLA4*, *TNFAIP3* and *CD40* loci); the remaining 21 confirmed alleles obtained $P_{GWAS} \leq 0.002$. We found modest evidence for association at *PADI4* (rs2240340, $P_{GWAS} = 0.01$, odds ratio (OR) = 1.06) and *FCRL3* (rs3761959, $P_{GWAS} = 0.001$, OR = 1.08), but not at *CD244* (rs3753389, $P_{GWAS} = 0.26$, OR = 1.03); SNPs at these three gene loci have been previously implicated in rheumatoid arthritis cases of Asian ancestry[3,20,21] (**Supplementary Table 1**).

We used three criteria to select 34 independent SNPs for replication after removing previously confirmed rheumatoid arthritis risk alleles. First, we selected all 11 SNPs with $P_{GWAS} < 10^{-6}$. Second, we selected nine SNPs (of the total 116 SNPs tested) with $P_{GWAS} < 0.0001$ and having been identified as functionally related to known rheumatoid arthritis

© 2010 Nature America, Inc. All rights reserved.

Table 2 Previously known SNPs associated with rheumatoid arthritis risk in Europeans

SNP			GWAS meta-analysis				Power	
Locus	ID	Gene(s)	Minor allele	MAF	OR (95% CI)	P_{GWAS}	$\alpha = 10^{-6}$	$\alpha = 5 \times 10^{-8}$
1p36	rs3890745*	*TNFRSF14*	C	0.32	0.89 (0.85–0.94)	3.6×10^{-6}	0.45	0.24
1p13	rs2476601	*PTPN22*	A	0.10	1.94 (1.81–2.08)	9.1×10^{-74}	1.00	1.00
1p13	rs11586238	*CD2, CD58*	G	0.24	1.13 (1.07–1.19)	1.0×10^{-5}	0.46	0.26
1q23	rs12746613*	*FCGR2A*	T	0.12	1.13 (1.06–1.21)	0.0004	0.11	0.04
1q31	rs10919563*	*PTPRC*	A	0.13	0.88 (0.82–0.94)	0.0002	0.10	0.03
2p16	rs13031237	*REL*	T	0.37	1.13 (1.07–1.18)	7.9×10^{-7}	0.67	0.45
2q11	rs10865035*	*AFF3*	A	0.47	1.12 (1.07–1.17)	2.0×10^{-6}	0.55	0.33
2q32	rs7574865	*STAT4*	T	0.22	1.16 (1.10–1.23)	2.9×10^{-7}	0.77	0.58
2q33	rs1980422	*CD28*	C	0.24	1.12 (1.06–1.18)	5.2×10^{-5}	0.32	0.15
2q33	rs3087243	*CTLA4*	A	0.44	0.87 (0.83–0.91)	1.2×10^{-8}	0.89	0.74
4q27	rs6822844	*IL2, IL21*	T	0.18	0.90 (0.84–0.95)	0.0007	0.08	0.02
6p21	rs6910071	*HLA-DRB1* (*0401 tag)	G	0.22	2.88 (2.73–3.03)	$<10^{-299}$	1.00	1.00
6q21	rs548234	*PRDM1*	C	0.33	1.10 (1.05–1.16)	9.7×10^{-5}	0.21	0.08
6q23	rs10499194	*TNFAIP3*	T	0.27	0.91 (0.87–0.96)	0.0007	0.11	0.03
6q23	rs6920220	*TNFAIP3*	A	0.22	1.22 (1.16–1.29)	8.9×10^{-13}	1.00	0.99
6q23	rs5029937	*TNFAIP3*	T	0.04	1.40 (1.24–1.58)	7.5×10^{-8}	0.95	0.86
6q25	rs394581*	*TAGAP*	C	0.30	0.91 (0.87–0.96)	0.0006	0.13	0.05
8p23	rs2736340	*BLK*	T	0.25	1.12 (1.07–1.18)	1.5×10^{-5}	0.34	0.17
9p13	rs2812378*	*CCL21*	G	0.34	1.10 (1.05–1.16)	0.0001	0.21	0.09
9q33	rs3761847	*TRAF1, C5*	G	0.43	1.13 (1.08–1.18)	2.1×10^{-7}	0.70	0.49
10p15	rs2104286	*IL2RA*	C	0.27	0.92 (0.87–0.97)	0.002	0.05	0.02
10p15	rs4750316	*PRKCQ*	C	0.19	0.87 (0.82–0.92)	2.0×10^{-6}	0.41	0.21
11p12	rs540386*	*TRAF6*	T	0.14	0.88 (0.83–0.94)	0.0003	0.13	0.05
12q13	rs1678542*	*KIF5A, PIP4K2C*	G	0.38	0.91 (0.87–0.96)	0.0002	0.20	0.08
20q13	rs4810485	*CD40*	T	0.25	0.85 (0.80–0.90)	2.8×10^{-9}	0.88	0.73
22q12	rs3218253*	*IL2RB*	A	0.26	1.09 (1.03–1.15)	0.002	0.07	0.02

GWAS meta-analysis results for previously known SNPs associated with rheumatoid arthritis risk among European populations. Listed are the chromosome, SNP ID and candidate gene(s) in the region. The minor allele and frequency (MAF) (positive strand in HapMap release 22, frequency in controls subjects), odds ratio (95% CI) and association *P* value are derived from our GWAS meta-analysis. Most SNPs have achieved $P < 5 \times 10^{-8}$ in combined analysis from previous studies; those with an asterisk (*) have been validated by replication in independent samples but may not have attained $P < 5 \times 10^{-8}$ in any single study. SNPs with strong evidence of association in Asians (including *PADI4*, rs2240340, $P_{GWAS} = 0.01$; *FCRL3*, rs3761959, $P_{GWAS} = 0.001$; and *CD244*, rs3753389, $P_{GWAS} = 0.26$) or suggestive evidence of association in Europeans are shown in **Supplementary Table 1**. Power was calculated based on the odds ratios from our meta-analysis at $\alpha = 10^{-6}$ (a threshold for selecting SNPs for replication in the current study) and $\alpha < 5 \times 10^{-8}$ (the threshold for genome-wide significance). Details of the power calculations can be found in the **Supplementary Note**.

risk loci by GRAIL[22], a bioinformatic analysis tool that identifies connections among genes in published abstracts (**Supplementary Fig. 6**). Third, we selected 14 SNPs (of the total 104 SNPs tested) with $P_{GWAS} < 0.01$ and previously known to be associated with other autoimmune diseases[23–31], as we found evidence for enrichment of autoimmune-associated SNPs in our rheumatoid arthritis GWAS (**Supplementary Fig. 7**). See **Supplementary Note** for additional details about our selection of SNPs for replication.

These 34 SNPs were genotyped in an independent set of 6,768 autoantibody-positive rheumatoid arthritis cases and 8,806 matched controls of European ancestry (**Table 1**). As in our GWAS, all cases were seropositive for disease-specific autoantibodies (anti-CCP or rheumatoid factor). For each SNP, we tested for replication of the GWAS meta-analysis association by calculating a one-tailed *P* value ($P_{replication}$) for the same allele in both analyses. Additionally, we conducted overall association tests across all 41,282 samples (GWAS meta-analysis plus the replication samples, $P_{overall}$) and considered $P_{overall} < 5 \times 10^{-8}$ to be a reproducible level of significance for rheumatoid arthritis risk association.

Of the 34 SNPs tested, 12 obtained a Bonferroni-corrected $P_{replication} < 0.0015$ (calculated as 0.05/34 total tests), and an additional nine SNPs obtained $P_{replication} < 0.05$. Ten out of these 34 SNPs achieved genome-wide significance in the combined analysis ($P_{overall} < 5 \times 10^{-8}$), indicating that these are validated rheumatoid arthritis risk alleles (**Table 3**). The other 11 SNPs that achieved replication significance had $P_{replication} < 0.05$ but had $P_{overall} > 5 \times 10^{-8}$, indicating highly suggestive but not genome-wide significant association (**Table 4**). Results for all 34 SNPs in the replication and combined analyses are shown in **Supplementary Table 2**.

Three of the ten newly validated rheumatoid arthritis risk loci have not been implicated in any previous genetic studies of rheumatoid arthritis or in studies of other autoimmune diseases. These SNPs are located at chromosome 2p14 (rs934734, $P_{overall} = 5.3 \times 10^{-10}$), 5q11 (rs6859219, $P_{overall} = 9.6 \times 10^{-12}$) and 5q21 (rs26232, $P_{overall} = 4.1 \times 10^{-8}$). Replication sample ORs for the minor alleles of these SNPs were 1.13, 0.85 and 0.93, respectively (**Table 3**).

Although no single gene can be declared causal as a result of this analysis, we labeled the new rheumatoid arthritis risk loci with the names of the most compelling candidate gene(s) from each region of linkage disequilibrium (LD) based upon GRAIL analysis and/or knowledge of rheumatoid arthritis pathogenesis[2,22]. At 2p14, the most significant SNP (rs934734) is located within intron 1 of *SPRED2*, encoding the sprouty-related, EVH1 domain–containing protein 2 (**Fig. 1a**), which has been shown to regulate $CD45^+$ hematopoietic cells via the Ras-MAP kinase pathway[32]. At 5q11, the most significant SNP (rs6859219) is located in *ANKRD55*, an ankyrin repeat domain–containing gene of unknown function (**Fig. 1b**). A more compelling immunological candidate, *IL6ST*, encoding interleukin 6 signal transducer, lies ~150 kb proximal to rs6859219 but outside the region of LD with associated SNPs. The *IL6ST* protein product, gp130, functions as a part of the receptor complex for the inflammatory cytokine IL6 (ref. 33). At 5q21, there is no obvious biological candidate gene; rs26232 lies within the intron of the predicted gene *C5orf30* (**Fig. 1c**).

Our study provides the first convincing evidence that four loci implicated in other autoimmune diseases are also associated with risk of rheumatoid arthritis. Of these, three of the four SNPs were selected for replication based on obtaining $P_{GWAS} < 10^{-6}$, regardless of their previously reported associations with autoimmune disease. These three SNPs are located at chromosome 3p14 (rs13315591, near *PXK*, $P_{overall} = 4.6 \times 10^{-8}$), 4p15 (rs874040, near *RBPJ*, $P_{overall} = 1.0 \times 10^{-16}$) and 6q27 (rs3093023, in *CCR6*, $P_{overall} = 1.5 \times 10^{-11}$). The 3p14 SNP lies 187 kb proximal to a SNP associated with systemic lupus erythematosus (SLE)[25]; the SLE SNP, intronic in the *PXK* gene, is only weakly associated with rheumatoid arthritis risk in our study (rs6445975, $P_{GWAS} = 0.03$, $r^2 = 0.15$ and $D' = 0.75$ with rs13315591). The 4p15 *RBPJ* SNP is in complete LD ($r^2 = 1$) with a SNP associated with risk of type 1 diabetes (T1D-associated SNP, rs10517086)[23], and

© 2010 Nature America, Inc. All rights reserved.

Table 3 Newly validated rheumatoid arthritis risk alleles

SNP						GWAS meta-analysis				Replication				Combined	
				Allele				MAF				MAF			
ID	Chr.	Position	Gene(s)	Major	Minor	P_{GWAS}	OR	Case	Control	$P_{replication}$	OR (95% CI)	Case	Control	$P_{overall}$	Cochran Q P
New rheumatoid arthritis and new autoimmune risk loci															
rs934734	2p14	65,507,237	*SPRED2*	A	G	3.2×10^{-7}	1.13	0.52	0.49	0.0002	1.13 (1.06–1.21)	0.53	0.51	5.3×10^{-10}	0.95
rs6859219	5q11	55,474,337	*ANKRD55, IL6ST*	C	A	2.5×10^{-9}	0.78	0.18	0.21	0.0002	0.85 (0.78–0.93)	0.19	0.22	9.6×10^{-12}	0.19
rs26232	5q21	102,624,619	*C5orf30*	C	T	4.3×10^{-7}	0.88	0.29	0.32	0.004	0.93 (0.88–0.98)	0.30	0.32	4.1×10^{-8}	0.82
New rheumatoid arthritis loci previously implicated in other autoimmune diseases[a]															
rs13315591	3p14	58,531,881	*PXK*	T	C	3.7×10^{-7}	1.29	0.10	0.09	0.002	1.13 (1.04–1.23)	0.09	0.08	4.6×10^{-8}	0.12
rs874040	4p15	25,784,466	*RBPJ*	G	C	1.9×10^{-7}	1.14	0.33	0.30	3.0×10^{-11}	1.18 (1.12–1.24)	0.34	0.30	1.0×10^{-16}	0.57
rs3093023	6q27	167,504,701	*CCR6*	G	A	3.3×10^{-7}	1.13	0.47	0.43	4.5×10^{-6}	1.11 (1.06–1.16)	0.46	0.43	1.5×10^{-11}	0.42
rs10488631	7q32	128,188,134	*IRF5*	T	C	2.8×10^{-6}	1.19	0.13	0.11	1.2×10^{-6}	1.25 (1.14–1.37)	0.13	0.10	4.2×10^{-11}	0.60
Associations at known rheumatoid arthritis risk loci[b]															
rs11676922	2q11	100,265,458	*AFF3*	A	T	6.9×10^{-7}	1.12	0.49	0.46	1.1×10^{-9}	1.15 (1.10–1.20)	0.48	0.45	1.0×10^{-14}	0.50
rs951005	9p13	34,733,681	*CCL21*	A	G	5.4×10^{-7}	0.84	0.14	0.16	6.7×10^{-5}	0.87 (0.81–0.93)	0.13	0.15	3.9×10^{-10}	0.61
rs706778	10p15	6,138,955	*IL2RA*	C	T	7.9×10^{-8}	1.14	0.44	0.40	1.5×10^{-5}	1.11 (1.06–1.17)	0.43	0.40	1.4×10^{-11}	0.36

Shown are GWAS, replication and combined meta-analysis results for SNPs that achieved genome-wide significance for association with rheumatoid arthritis risk. Listed are the rs ID for each SNP, the chromosomal/cytological band (Chr.), position in human genome build 36, candidate genes in the region (selected by GRAIL analysis or manually based on immunological function; see text and **Supplementary Note**), and major and minor alleles (positive strand in HapMap release 22, major/minor based on frequency in GWAS controls). The association *P* value, odds ratio (OR) with respect to minor allele (95% CI for replication analysis) and minor allele frequencies (MAF) in cases and controls are listed for our GWAS and replication analyses. For the combined analysis, the overall association *P* value and the *P* value for Cochran's Q test for heterogeneity are listed.
[a]Implicated in other autoimmune diseases: *PXK* is associated with SLE (rs6445975, $r^2 = 0.15$ with rs13315591); *RBPJ* is associated with T1D (rs10517086, $r^2 = 1$ with rs874040); *CCR6* is associated with Crohn's disease (rs2301436, $r^2 = 0.48$ with rs3093023); and *IRF5* SNP rs10488631 is associated with SLE. [b]Previously implicated in rheumatoid arthritis: *AFF3* SNP rs11676922 has recently been reported[14]; *CCL21* SNP rs2812378 ($r^2 = 0.06$ with rs951005); and *IL2RA* SNP rs2104286 ($r^2 = 0.25$ with rs706778).

© 2010 Nature America, Inc. All rights reserved.

the same allele confers risk in both diseases. The 6q27 *CCR6* SNP rs3093023 is in LD with a SNP associated with Crohn's disease[24], which is only weakly associated with rheumatoid arthritis risk in our study (Crohn's-associated SNP rs2301436, $P_{GWAS} = 0.045$, $r^2 = 0.48$ and $D' = 0.80$).

The *IRF5* SNP rs10488631 ($P_{GWAS} = 2.8 \times 10^{-6}$), chosen because of its association with SLE[34,35], was convincingly associated with autoantibody-positive rheumatoid arthritis in our study ($P_{replication} = 1.2 \times 10^{-6}$, $P_{overall} = 4.2 \times 10^{-11}$). A previous study[36] proposed that the complex SLE risk associations in *IRF5* are explained by three independent groups of SNPs, each consisting of SNPs in tight LD with each other. In our dataset, these groups are represented by rs10488631 (group 1), rs729302 (group 2) and rs4728142 (group 3). In addition to finding association with the group 1 SNP rs10488631, we found evidence for association with group 3 SNP rs4728142 ($P_{GWAS} = 7.2 \times 10^{-6}$), which is in LD with a variant that alters *IRF5* polyadenylation and expression[36]. However, we found no evidence for association of the representative group 2 SNP (rs729302, $P_{GWAS} = 0.15$). Conditional analyses indicated that the group 1 and group 3 effects are independent of each other (**Supplementary Table 3**). We note that the group 3 SNP rs3807306 has been suggested to be associated with autoantibody-negative rheumatoid arthritis[37]; however,

Table 4 Suggestive rheumatoid arthritis risk alleles

SNP						GWAS meta-analysis				Replication				Combined	
				Allele				MAF				MAF			
ID	Chr.	Pos (HG18)	Gene(s)	Major	Minor	P_{GWAS}	OR	Case	Control	$P_{replication}$	OR (95% CI)	Case	Control	$P_{overall}$	Cochran Q P
rs7543174[a]	1q21	151,340,745	*IL6R*	T	C	7.9×10^{-5}	1.13	0.18	0.16	0.01	1.07 (1.01–1.13)	0.19	0.18	1.2×10^{-5}	0.06
rs840016[a]	1q24	164,140,328	*CD247*[c]	C	T	3.6×10^{-5}	0.90	0.39	0.42	0.006	0.92 (0.86–0.98)	0.38	0.40	1.6×10^{-6}	0.62
rs13119723[b]	4q27	123,575,918	*IL2, IL21*[c]	A	G	0.001	0.89	0.13	0.15	6.7×10^{-5}	0.87 (0.81–0.93)	0.15	0.17	6.8×10^{-7}	0.46
rs11594656[b]	10p15	6,162,015	*IL2RA*[c]	T	A	0.0002	0.90	0.23	0.25	0.04	0.95 (0.90–1.00)	0.24	0.25	0.0001	0.86
rs2793108[b]	10p11	31,419,111	*ZEB1*	T	C	0.002	0.93	0.40	0.43	0.001	0.93 (0.89–0.98)	0.41	0.43	1.4×10^{-5}	0.73
rs3184504[b]	12q24	110,347,328	*SH2B3*	T	C	0.004	0.93	0.49	0.49	0.0002	0.92 (0.88–0.96)	0.48	0.49	6.0×10^{-6}	0.08
rs7155603[a]	14q24	75,030,289	*BATF*[c]	A	G	1.0×10^{-5}	1.16	0.21	0.19	0.001	1.12 (1.04–1.20)	0.23	0.21	1.1×10^{-7}	0.74
rs8045689[a]	16p11	28,895,770	*CD19, NFATC2IP*[c]	T	C	5.3×10^{-5}	1.14	0.32	0.30	0.01	1.06 (1.01–1.12)	0.32	0.30	2.4×10^{-5}	0.35
rs2872507[b]	17q12	35,294,289	*IKZF3*[c]	G	A	4.7×10^{-5}	1.10	0.49	0.47	0.002	1.08 (1.02–1.14)	0.48	0.46	9.4×10^{-7}	0.69
rs11203203[b]	21q22	42,709,255	*UBASH3A*[c]	G	A	2.5×10^{-5}	1.11	0.39	0.37	0.02	1.07 (1.00–1.14)	0.38	0.37	3.8×10^{-6}	0.49
rs5754217[b]	22q11	20,264,229	*UBE2L3*	G	T	0.0007	1.10	0.22	0.19	0.01	1.07 (1.01–1.13)	0.22	0.21	4.8×10^{-5}	0.79

GWAS, replication and combined meta-analysis results for SNPs with highly suggestive associations with rheumatoid arthritis risk (defined as $P < 0.05$ in our replication samples). As in **Table 2**, listed for each SNP are the rs ID, chromosomal location (Chr.) and position, candidate gene(s), major and minor alleles, GWAS and replication analysis *P* value, OR and case and control minor allele frequencies, and the combined analysis *P* values for association and for the Cochran's *Q* test for heterogeneity.
[a]Selected for replication based on GRAIL-based P_{text} score. [b]Selected for replication based on $P_{GWAS} < 0.01$ and a validated autoimmune disease association: rs13119723, *IL2-IL21* for rheumatoid arthritis and celiac disease; rs11594656, *IL2RA* for T1D and MS; rs2793108, *ZEB1*, T1D; rs3184504, *SH2B3*, celiac disease and T1D; rs2872507, *IKZF3*, Crohn's disease; rs11203203, *UBASH3A* for T1D; rs5754217, *UBE2L3* for SLE; the *ZEB1* SNP was from the May 2009 release of the online T1D database (see URLs). [c]Autoimmune disease associations for SNPs other than those tested for replication in the present study (see main text, **Fig. 2** and the **Supplementary Note**): *CD247* for Crohn's disease; *IL2-IL21* for celiac disease, rheumatoid arthritis and T1D; *IL2RA* for MS, rheumatoid arthritis and T1D; *BATF* for T1D; *CD19-NFATC2IP* for T1D; *IKZF3* for T1D; and *UBASH3A* for Crohn's disease and T1D.

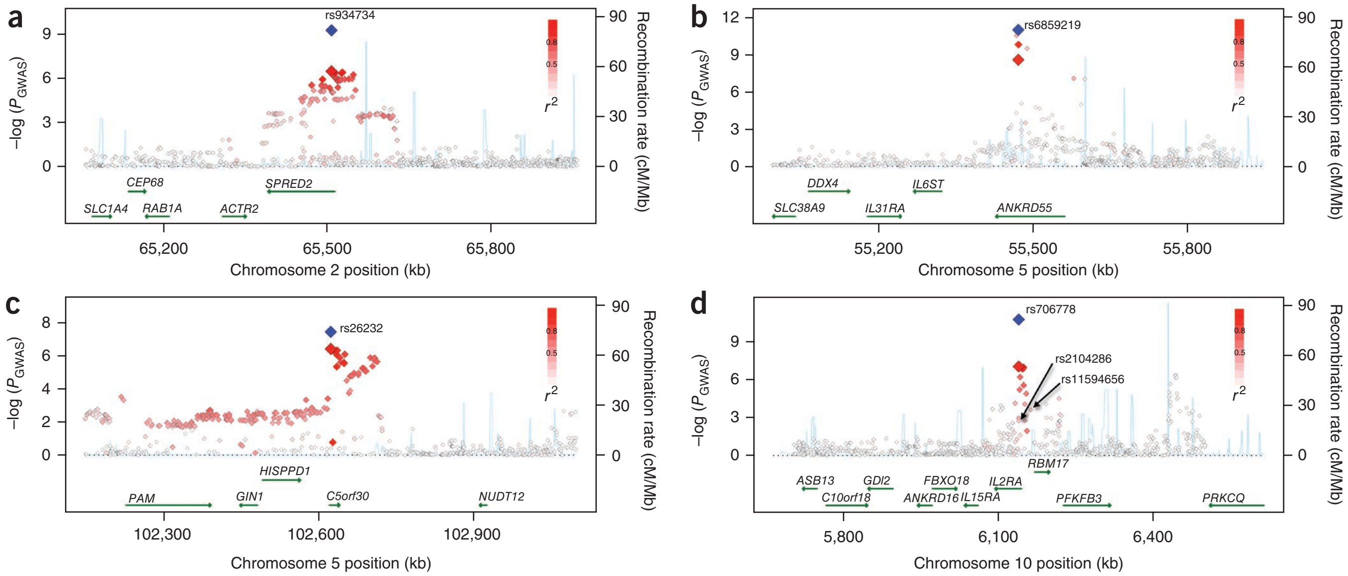


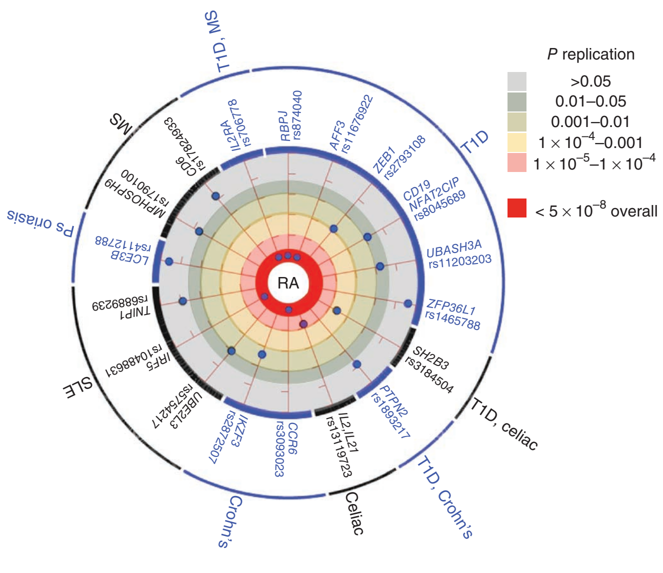
Figure 1 Associations with rheumatoid arthritis risk across four loci. Regional association plots show strength of association ($-\log_{10}(P)$) versus chromosomal position (kb) for all SNPs across 1 Mb regions centered on the newly validated SNPs (labeled). P_{GWAS} values are plotted with diamonds for all SNPs, shaded white to red by the degree of LD (r^2; see inset) with the validated SNP (larger red diamond). $P_{overall}$ in combined analysis of GWAS and replication collections is plotted with a blue diamond. Local recombination rates estimated from HapMap CEU (cM/Mb, blue line) are plotted against the secondary *y* axis, showing recombination hotspots across the region. Labeled green arrows below the plots indicate genes and their orientations. (**a**) 2p14, *SPRED2* locus. (**b**) 5q11, *IL6ST-ANKRD55* locus. (**c**) 5q21, *C5orf30* locus. (**d**) 10p15, *IL2RA* locus.

our results suggest that this SNP is associated with autoantibody-positive rheumatoid arthritis ($P_{GWAS} = 4.5 \times 10^{-6}$).

Our GWAS meta-analysis identified new alleles exhibiting independent effects from previously implicated rheumatoid arthritis risk alleles at two other loci, *CCL21* (rs951005, $P_{overall} = 3.9 \times 10^{-10}$)[11] and *IL2RA* (rs706778, $P_{overall} = 3.3 \times 10^{-11}$)[15]. Our new *CCL21* SNP has little LD with the previous rheumatoid arthritis risk-associated SNP rs2812378 ($P_{GWAS} = 1.6 \times 10^{-7}$, $r^2 = 0.06$ and $D' = 0.83$)[11]. Conditional analyses indicated that the associations of these two SNPs are partly independent (**Supplementary Table 3**); the new SNP rs951005 remains significant ($P = 1.7 \times 10^{-5}$) conditional on the previously known SNP rs2812378 and conditional on rs951005, the previously known rs2812378 remains nominally significant ($P = 0.01$).

SNPs at *IL2RA* are known to be associated with multiple autoimmune diseases, including rheumatoid arthritis[15], T1D[23] and multiple sclerosis (MS)[38,39]. The *IL2RA* SNP identified in our GWAS (rs706778, $P_{overall} = 3.3 \times 10^{-11}$; **Fig. 1d**) is in partial LD with a SNP previously implicated in rheumatoid arthritis, T1D and MS (rs2104286, $r^2 = 0.25$ and $D' = 1$)[15], as well as a SNP previously implicated in T1D and MS (rs11594656, $r^2 = 0.24$ and $D' = 1$)[23]. In fact, T1D risk is conferred by a haplotype tagged by rs2104286[T]-rs11594656[T][38]. In our GWAS dataset, rs706778 correlated strongly with the T1D risk haplotype ($r^2 = 0.83$), and conditional and haplotype analyses demonstrated that risk of rheumatoid arthritis is better explained by this haplotype (rs2104286[T]-rs11594656[T]) than by rs706778 (**Supplementary Tables 3** and **4**).

Figure 2 Previously validated autoimmune SNPs tested in our replication study. Eighteen SNPs tested in our replication samples were in LD (defined as $r^2 > 0.3$) with a validated autoimmune risk allele. Of these, five were validated as rheumatoid arthritis risk alleles in our study ($P_{overall} < 5 \times 10^{-8}$, inner most circle), six were suggestive associations ($P_{replication} < 0.05$ but $P_{overall} > 5 \times 10^{-8}$) and six demonstrated no evidence of association in our replication samples ($P_{replication} \geq 0.05$). For the 12 SNPs with suggestive or no evidence of association, each SNP is plotted by the strength of association with rheumatoid arthritis risk in the replication samples; those closer to the inner circle have more significant $P_{replication}$. All of the rheumatoid arthritis risk alleles confer risk in the same direction as the validated autoimmune risk alleles (when the same allele or a near perfect proxy was tested). We include the following as 'autoimmune' diseases in our study, listed on the outside of the circle, although these reflect diseases along the autoimmune-inflammatory spectrum: systemic lupus erythematosus (SLE), celiac disease, Crohn's disease, multiple sclerosis (MS), psoriasis, and type 1 diabetes (T1D) but other autoimmune diseases are not included (for example, autoimmune thyroiditis). Note that there are SNPs associated with rheumatoid arthritis and other autoimmune diseases not shown; we only include those SNPs tested as part of the current study. See Online Methods for details about the SNPs validated in other autoimmune diseases.



© 2010 Nature America, Inc. All rights reserved.

A SNP at the *AFF3* locus has been previously implicated in rheumatoid arthritis[14], with equivocal evidence for association with T1D[23,40]. Our study provides strong evidence that this locus is associated with risk of autoantibody-positive rheumatoid arthritis (rs11676922, $P_{overall} = 1.0 \times 10^{-14}$, OR = 1.14).

Eighteen of the 34 SNPs that we tested in replication (based on the criteria above) are in LD with previously validated autoimmune risk alleles (**Fig. 2**). Of these, five SNPs demonstrated genome-wide significance ($P_{overall} < 5 \times 10^{-8}$) and seven showed suggestive evidence of association ($P_{replication} < 0.05$) in our study; only six SNPs showed no evidence of association in the replication stage. Though not meant as a complete comparison of all known rheumatoid arthritis and other autoimmune disease risk alleles, these results underscore the emerging overlap in the genetic bases of rheumatoid arthritis and other autoimmune diseases[41,42]. Although additional replication in large sample collections is required to confirm the suggestive associations, many of the SNPs found here likely represent true rheumatoid arthritis risk alleles.

Several of the SNP associations seen here further implicate T cells in rheumatoid arthritis pathogenesis. *RBPJ* (encoding recombination site binding protein J and also known as *CSL*) encodes a transcription factor within the Notch signaling pathway. Rbpj-deficient mice have no T-cell development, whereas early B-cell development is maintained[43]. CCR6 is a cell surface protein that distinguishes Th17 cells from other CD4$^+$ helper T cells[44]. Synoviocytes from arthritic joints of mice and individuals with rheumatoid arthritis produce CCL20, a CCR6 ligand. Notably, anti-Ccr6 monoclonal antibodies substantially inhibit mouse arthritis, suggesting that CCR6 could be a therapeutic target in human rheumatoid arthritis[44]. *CD247* ($P_{replication} = 0.006$, $P_{overall} = 1.6 \times 10^{-6}$; **Table 3**) encodes the T-cell receptor zeta chain, a subunit of the T-cell receptor–CD3 complex that, when altered, causes an inflammatory arthritis in mouse[45,46]. The zeta chain plays an important role in coupling antigen recognition to several intracellular signal-transduction pathways. These results build upon knowledge of T-cell activation-differentiation in rheumatoid arthritis pathogenesis as evidenced by genetic associations at *HLA-DRB1* (encoding a protein that presents antigens to T cells), *PTPN22* (encoding a protein that alters T-cell thresholds), *CTLA4* (encoding a co-stimulator for T-cell activation), *IL2RA* (encoding a protein that mediates IL2-dependent T-cell responses) and *STAT4* (encoding a transcription factor involved in Th1 cell differentiation), among others.

In conclusion, we find convincing evidence for association with risk of autoantibody-positive rheumatoid arthritis at ten loci, seven of which represent new risk loci for rheumatoid arthritis. We estimate that the now >30 validated non-MHC rheumatoid arthritis risk alleles explain 3.9% of the total disease variance, with 0.67% of this variance being due to the new risk alleles reported here. It is clear that additional risk alleles remain to be identified, as current genetic discoveries explain only 16% of disease variance (including an estimated 12% for the MHC region[47]), whereas more than half the risk of autoantibody-positive rheumatoid arthritis is thought to be genetic[47,48]. The number of SNPs with suggestive evidence of association in our study (**Table 4**) further indicates that many more common risk alleles with modest effect sizes remain to be discovered. In addition to common variants, the roles of rare variants, copy number variants and epigenetic modifications will need to be explored with newer genomic technologies.

URLs. EIGENSTRAT software, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm; SNPTEST software, https://mathgen.stats.ox.ac.uk/impute/impute.html; GRAIL software, http://www.broadinstitute.org/mpg/grail/; GWAS meta-analysis results, http://www.broadinstitute.org/ftp/pub/rheumatoid_arthritis/Stahl_etal_2010NG/; T1D database, http://www.T1Dbase.org.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

R.M.P. is supported by grants from the US National Institutes of Health (NIH) (R01-AR057108, R01-AR056768 and U54 RR020278), a private donation from the Fox Trot Fund, the William Randolph Hearst Fund of Harvard University, the American College of Rheumatology 'Within Our Reach' campaign and a Career Award for Medical Scientists from the Burroughs Wellcome Fund. S.R. is supported by an NIH Career Development Award (1K08AR055688-01A1) and an American College of Rheumatology Bridge Grant. F.A.S.K. is supported by an EMBO-UNESCO L'Oreal Fellowship. The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278 from the National Center for Research Resources. The BRASS Registry is supported by a grant from Crescendo and Biogen-Idec. EIRA is supported by grants from the Swedish Medical Research council, the Swedish Council for Working Life and Social Research, King Gustaf V's 80-year foundation, the Swedish Rheumatism Foundation, Stockholm County Council, from Vinnova and the insurance company AFA. NARAC is supported by the NIH (NO1-AR-2-2263 and RO1 AR44422). L.A.C. is supported by the NIH (R01 AI065841 and 5-M01-RR-00079). The Nurses Health Study is supported by NIH grants P01 CA87969, CA49449, CA67262, CA50385, AR049880-06 and AR47782. This research was also supported in part by the Intramural Research Program of the National Institute of Arthritis, Musculoskeletal and Skin Diseases of the US National Institutes of Health. This research was also supported in part by grants to K.A.S. from the Canadian Institutes for Health Research (MOP79321 and IIN-84042) and the Ontario Research Fund (RE01061) and by a Canada Research Chair. Genotyping of United Kingdom Rheumatoid Arthritis Genetics samples was supported by the Arthritis Research campaign arc grant reference number 17552 and by the Manchester Biomedical Research Centre and Manchester Academy of Health Sciences. C.W. was funded by the Netherlands Organization for Scientific Research (VICI grant 918.66.620). We acknowledge the help of B.A.C. Dijkmans, D. van Schaardenburg, A. Salvador Peña, P.L. Klarenbeek, Z. Zhang, M.T. Nurmohamed, W.F. Lems, R.R. J. van de Stadt, W.H. Bos, J. Ursum, M.G.M. Bartelds, D.M. Gerlag, M.G.H. van der Sande, C.A. Wijbrandts and M.M.J. Herenius in gathering Genetics Network Rheumatology Amsterdam subject samples and data. We thank the Myocardial Infarction Genetics Consortium (MIGen) study for the use of genotype data from their healthy controls in our study. The MIGen study was funded by the US National Institutes of Health and National Heart, Lung, and Blood Institute's SNP Typing for Association with Multiple Phenotypes from Existing Epidemiological Data (STAMPEED) genomics research program R01HL087676 and a grant from the National Center for Research Resources. We thank J. Seddon, Progression of AMD Study, Age-Related Macular Degeneration (AMD) Registry Study, Family Study of AMD, The US Twin Study of AMD and the Age-Related Eye Disease Study (AREDS) for use of genotype data from their healthy controls in our study. We thank D. Hafler and the Multiple Sclerosis Collaborative for use of genotype data from their healthy controls recruited at Brigham and Women's Hospital.

AUTHOR CONTRIBUTIONS

Study design: R.M.P., S.R., E.A.S. **Analysis:** E.A.S. (lead), S.R., F.A.S.K., R.C. **Sample procurement and data generation:** J.W., K.A.S., P.K.G., L.K., N.A.S., M.E.W., C.W., M.J.H.C., N.d.V., P.P.T., E.W.K., R.E.M.T., T.W.J.H., A.B.B. (leads); E.F.R., G.X., S.E., B.P.T., Y.L., A.Z., A.H., C.G., L.A., C.I.A., K.G.A., A.B., J.B., E.B., N.P.B., J.J.C., J. Coblyn, K.H.C., L.A.C., J.B.A.C., J. Cui, P.I.W.d.B., P.L.D.J., B.D., P.E., E.F., P.H., L.J.H., D.L.K., X.K., A.T.L., X.L., P.M., A.W.M., L.P., M.D.P., T.R.D.J.R., D.M.R., M.S., M.F.S., S.S., W.T., A.H.M.v.d.H.-v.M., I.E.v.d.H.-B., C.E.v.d.S., P.L.C.M.v.R., A.G.W., G.J.W., B.P.W., BIRAC and YEAR consortia. **Writing:** R.M.P., E.A.S. (leads); S.R., F.A.S.K. (primary contributors); J.W., K.A.S., P.K.G., L.K., N.A.S., M.E.W., C.W., M.J.H.C., N.d.V., P.P.T., E.W.K., R.E.M.T., T.W.J.H., A.B.B., E.F.R., G.X., S.E., B.P.T., Y.L., A.Z., A.H., C.G., L.A., C.I.A., K.G.A., A.B., J.B., E.B., N.P.B., J.J.C., J. Coblyn, K.H.C., L.A.C., J.B.A.C., J. Cui, P.I.W.d.B., P.L.D.J., B.D., P.E., E.F., P.H., L.J.H., D.L.K., X.K., A.T.L., X.L., P.M., A.W.M., L.P., M.D.P., T.R.D.J.R., D.M.R., M.S., M.F.S., S.S., W.T., A.H.M.v.d.H.-v.M., I.E.v.d.H.-B., C.E.v.d.S., P.L.C.M.v.R., A.G.W., G.J.W., B.P.W.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

© 2010 Nature America, Inc. All rights reserved.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Silman, A.J. & Pearson, J.E. Epidemiology and genetics of rheumatoid arthritis. *Arthritis Res.* **4** Suppl 3, S265–S272 (2002).
2. Klareskog, L., Catrina, A.I. & Paget, S. Rheumatoid arthritis. *Lancet* **373**, 659–672 (2009).
3. Suzuki, A. *et al.* Functional haplotypes of *PADI4*, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
4. Begovich, A.B. *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.* **75**, 330–337 (2004).
5. Kurreeman, F.A. *et al.* A candidate gene approach identifies the *TRAF1*/*C5* region as a risk factor for rheumatoid arthritis. *PLoS Med.* **4**, e278 (2007).
6. Plenge, R.M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.* **39**, 1477–1482 (2007).
7. Plenge, R.M. *et al.* *TRAF1-C5* as a risk locus for rheumatoid arthritis—a genomewide study. *N. Engl. J. Med.* **357**, 1199–1209 (2007).
8. Remmers, E.F. *et al.* *STAT4* and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.* **357**, 977–986 (2007).
9. Thomson, W. *et al.* Rheumatoid arthritis association at 6q23. *Nat. Genet.* **39**, 1431–1433 (2007).
10. Zhernakova, A. *et al.* Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. *Am. J. Hum. Genet.* **81**, 1284–1288 (2007).
11. Raychaudhuri, S. *et al.* Common variants at *CD40* and other loci confer risk of rheumatoid arthritis. *Nat. Genet.* **40**, 1216–1223 (2008).
12. Raychaudhuri, S. *et al.* Genetic variants at *CD28*, *PRDM1* and *CD2*/*CD58* are associated with rheumatoid arthritis risk. *Nat. Genet.* **41**, 1313–1318 (2009).
13. Gregersen, P.K. *et al.* *REL*, encoding a member of the NF-κB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* **41**, 820–823 (2009).
14. Barton, A. *et al.* Identification of AF4/FMR2 family, member 3 (*AFF3*) as a novel rheumatoid arthritis susceptibility locus and confirmation of two further pan-autoimmune susceptibility genes. *Hum. Mol. Genet.* **18**, 2518–2522 (2009).
15. Barton, A. *et al.* Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat. Genet.* **40**, 1156–1159 (2008).
16. 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.* **39**, 906–913 (2007).
17. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
18. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
19. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
20. Suzuki, A. *et al.* Functional SNPs in *CD244* increase the risk of rheumatoid arthritis in a Japanese population. *Nat. Genet.* **40**, 1224–1229 (2008).
21. Kochi, Y. *et al.* A functional variant in *FCRL3*, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. *Nat. Genet.* **37**, 478–485 (2005).
22. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
23. Barrett, J.C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).
24. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
25. Harley, J.B. *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in *ITGAM*, *PXK*, *KIAA1542* and other loci. *Nat. Genet.* **40**, 204–210 (2008).
26. Nair, R.P. *et al.* Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways. *Nat. Genet.* **41**, 199–204 (2009).
27. de Cid, R. *et al.* Deletion of the late cornified envelope *LCE3B* and *LCE3C* genes as a susceptibility factor for psoriasis. *Nat. Genet.* **41**, 211–215 (2009).
28. Hom, G. *et al.* Association of systemic lupus erythematosus with *C8orf13-BEK* and *ITGAM-ITGAX*. *N. Engl. J. Med.* **358**, 900–909 (2008).
29. van Heel, D.A. *et al.* A genome-wide association study for celiac disease identifies risk variants in the region harboring *IL2* and *IL21*. *Nat. Genet.* **39**, 827–829 (2007).
30. Hunt, K.A. *et al.* Newly identified genetic risk variants for celiac disease related to the immune response. *Nat. Genet.* **40**, 395–402 (2008).
31. De Jager, P.L. *et al.* Meta-analysis of genome scans and replication identify *CD6*, *IRF8* and *TNFRSF1A* as new multiple sclerosis susceptibility loci. *Nat. Genet.* **41**, 776–782 (2009).
32. Nobuhisa, I. *et al.* Spred-2 suppresses aorta-gonad-mesonephros hematopoiesis by inhibiting MAP kinase activation. *J. Exp. Med.* **199**, 737–742 (2004).
33. Ernst, M. & Jenkins, B.J. Acquiring signalling specificity from the cytokine receptor gp130. *Trends Genet.* **20**, 23–32 (2004).
34. Graham, R.R. *et al.* A common haplotype of interferon regulatory factor 5 (*IRF5*) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.* **38**, 550–555 (2006).
35. Sigurdsson, S. *et al.* Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am. J. Hum. Genet.* **76**, 528–537 (2005).
36. Graham, R.R. *et al.* Three functional variants of IFN regulatory factor 5 (*IRF5*) define risk and protective haplotypes for human lupus. *Proc. Natl. Acad. Sci. USA* **104**, 6758–6763 (2007).
37. Sigurdsson, S. *et al.* Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. *Arthritis Rheum.* **56**, 2202–2210 (2007).
38. Dendrou, C.A. *et al.* Cell-specific protein phenotypes for the autoimmune locus *IL2RA* using a genotype-selectable human bioresource. *Nat. Genet.* **41**, 1011–1015 (2009).
39. Maier, L.M. *et al.* *IL2RA* genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet.* **5**, e1000322 (2009).
40. Todd, J.A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* **39**, 857–864 (2007).
41. Zhernakova, A., van Diemen, C.C. & Wijmenga, C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat. Rev. Genet.* **10**, 43–55 (2009).
42. Plenge, R.M. Recent progress in rheumatoid arthritis genetics: one step towards improved patient care. *Curr. Opin. Rheumatol.* **21**, 262–271 (2009).
43. Rothenberg, E.V., Moore, J.E. & Yui, M.A. Launching the T-cell-lineage developmental programme. *Nat. Rev. Immunol.* **8**, 9–21 (2008).
44. Hirota, K. *et al.* Preferential recruitment of CCR6-expressing Th17 cells to inflamed joints via CCL20 in rheumatoid arthritis and its animal model. *J. Exp. Med.* **204**, 2803–2812 (2007).
45. Wucherpfennig, K.W., Call, M.J., Deng, L. & Mariuzza, R. Structural alterations in peptide-MHC recognition by self-reactive T cell receptors. *Curr. Opin. Immunol.* **21**, 590–595 (2009).
46. Sakaguchi, N. *et al.* Altered thymic T-cell selection due to a mutation of the ZAP-70 gene causes autoimmune arthritis in mice. *Nature* **426**, 454–460 (2003).
47. MacGregor, A.J. *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* **43**, 30–37 (2000).
48. van der Woude, D. *et al.* Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum.* **60**, 916–923 (2009).

© 2010 Nature America, Inc. All rights reserved.

ONLINE METHODS

Sample collections. Case-control collections are listed in **Table 1** and described in detail in our previous studies[6,7,11,13,19]. Collections were composed entirely of individuals of self-described European ancestry, and all cases either met the 1987 American College of Rheumatology criteria for diagnosis of rheumatoid arthritis[49] or were diagnosed by board-certified rheumatologists. Rheumatoid arthritis cases were further limited to only individuals who either were anti-CCP positive or, if anti-CCP status data were missing, were rheumatoid factor positive. The BRASS rheumatoid arthritis samples have been used in our previous 100K GWAS[6] but are here presented for the first time genotyped with the Affymetrix 6.0 array. In the current study, controls were matched to BRASS rheumatoid arthritis cases using principal-components analysis from GWAS data from three separate studies: controls from a multiple sclerosis GWAS[31], controls from an age-related macular degeneration GWAS[50] and controls from a myocardial infarction GWAS[51]. Wellcome Trust Case Control Consortium (WTCCC) collection controls included the 1958 Birth and National Blood Service cohorts, as well as cases with non-autoimmune diseases (individuals with bipolar disorder, cardiovascular disease, hypertension and type 2 diabetes)[15]. All GWAS collections except those from the WTCCC were restricted to control subjects matching cases using principal-components analysis of GWAS data. All replication sample collections were composed of epidemiologically and/or geographically matched cases and controls, except the NARAC II collection, which was case-control matched based on genotypes at a set of ancestry-informative markers as previously described[11]. The eight replication samples included: (i) CCP-positive cases and controls from Halifax and Toronto (CANADA-II)[13]; (ii) rheumatoid factor–positive Dutch cases from Groningen and Nijmegen, together with geographically matched controls[52]; (iii) CCP-positive Dutch cases and controls collected from the greater Amsterdam region (GENRA)[53]; (iv) North American rheumatoid factor–positive cases and controls matched on gender, age and grandparental country of origin from the Genomics Collaborative Initiative (GCI)[4]; (v) CCP-positive or rheumatoid factor–positive Dutch cases and controls from Leiden University Medical Center (LUMC)[5]; (vi) CCP-positive cases drawn from North American clinics and controls from the New York Cancer Project (together this collection is called NARAC-II)[7]; (vii) CCP-positive or rheumatoid factor–positive cases recruited at multiple sites in the United Kingdom by the United Kingdom Rheumatoid Arthritis Genetics (UKRAG) collaboration[9]; and (viii) CCP-positive or rheumatoid factor–positive cases identified by chart review from the Nurses Health Study (NHS) and matched controls based on age, gender, menopausal status and hormone use[54]. We used available SNP data from this and previous studies to identify genetically identical samples from the same country[13]; we assumed these represented duplicated individuals and removed them. Institutional review boards at each collection site approved the study, and all individuals gave their informed consent.

Genotyping. The BRASS GWAS collection was genotyped on the Affymetrix GeneChip 6.0 platform at the Broad Institute (Boston, USA). All other GWAS collections were genotyped as previously described[7,13,19]. Genotype data for GWAS samples from rheumatoid arthritis and other disease studies were obtained with permission from the investigators and/or disease consortia. Additional shared control GWAS genotype data were obtained from the US National Institute of Mental Health through a formal application and approval process (part of the BRASS collection) and from the Illumina iControls database (NARAC III). For each GWAS collection, quality control was implemented in the cases and in each control cohort separately and then again in the merged collection data. Quality control steps included filtering SNPs and individuals with >5% missing data, followed by filtering SNPs with MAF <1% and a χ^2 test of Hardy-Weinberg equilibrium (HWE) $P_{\mathrm{HWE}} < 10^{-6}$. For the WTCCC collection, genotyped on the older Affymetrix 500K platform, we implemented more stringent quality control criteria (>1% missing data, MAF < 1% and $P_{\mathrm{HWE}} < 10^{-5}$). We then used individual-pairwise identity-by-state estimates to remove occasional related and potentially contaminated samples. Data processing and quality control filtering were performed in PLINK[55]. Additional details are described in the **Supplementary Note**.

The 34 SNPs chosen for replication, as well as proxy SNPs, were directly genotyped in each of eight collections (**Table 1** and **Supplementary Note**). Canadian samples were genotyped on the Sequenom iPlex platform at University of Toronto, Mount Sinai Hospital and University Health Network (Toronto, Canada); Dutch and GENRA samples were genotyped on the Sequenom iPlex platform at the Broad Institute (Cambridge, Massachusetts, USA); UKRAG samples were genotyped on the Sequenom iPlex platform at The University of Manchester (Manchester, UK); GCI and LUMC samples were genotyped by kinetic PCR at Celera (Alameda, California, USA); NARAC II samples were genotyped on the Sequenom iPlex platform at the NIH (Bethesda, Maryland, USA); and NHS samples were genotyped on the Biotrove OpenArray platform at the Channing Laboratory, Harvard Medical School (Boston, USA). Quality control exclusion criteria for SNPs in each replication or validation collection were 10% missing data, MAF < 1% and $P_{\mathrm{HWE}} < 10^{-3}$. If a given SNP failed in genotyping or quality control in a collection, a proxy SNP ($r^2 > 0.8$) with the least missing data (if available) was used instead. See **Supplementary Note** for details.

Genome-wide association analyses. To address population stratification and remove outliers in our GWAS for BRASS, Canada, EIRA, NARAC I and NARAC III, we performed principal-component analysis using EIGENSTRAT[56]. For BRASS, Canada, NARAC I and NARAC III, we further removed poorly matched controls based on case-control Euclidean distances calculated from five principal components[11]. Once matched, imputation was conducted on GWAS genotype data for each GWAS collection separately, using the IMPUTE software[16] and haplotype-phased HapMap Phase 2 European CEU founders as a reference panel. Imputation yielded posterior genotype probabilities as well as imputation quality scores at SNPs not genotyped with a MAF ≥1% in HapMap CEU.

We conducted logistic regression analysis for each SNP in each GWAS collection to estimate the regression coefficients (β) and the z-scores for allele counts (using an additive model), which were then genomic-control corrected[17]. λ_{GC} values for genotyped SNPs only as compared to the values for all SNPs together (**Supplementary Table 5**) verified that logistic regression controlled for any deflation in the distribution of association test results[18]. We then conducted a meta-analysis to combine results across datasets for 2,554,714 SNPs with high quality genotype data in one or more collections (see **Supplementary Note**) by summing inverse variance-weighted β and z-scores[18] and again genomic-control corrected our results. We also conducted Cochran's Q tests for heterogeneity across collections using the β coefficients for each collection for which results were available for a given SNP. Detailed descriptions of all analyses and results are provided in the **Supplementary Note**.

Replication analysis. Replication and combined analyses were conducted followed the GWAS meta-analysis; after matching in the NARAC II collection (**Table 1**)[11] and removing spurious duplicate samples[12], logistic regression was used to test for association, and inverse variance-weighted z-scores were summed across collections. Replication association tests were one-tailed for the same allele as being risk or protective as in the GWAS meta-analysis. Results of two alternative analyses to control for population stratification are reported in **Supplementary Table 6**.

For the 34 SNPs we tested in replication, we searched for SNPs in LD ($r^2 > 0.3$) that were validated in other autoimmune diseases (see main text and **Fig. 2**). The haplotype tagged by the *IL2RA* SNP, rs706778, is associated with T1D and MS[38,39]; the *SH2B3* SNP is associated with both T1D and Celiac disease[23,30]. The *CCR6* SNP rs3093023 is in partial LD with a SNP associated with Crohn's disease (rs2301436, $r^2 = 0.48$)[24]. The *AFF3* SNP has an equivocal association with T1D (where the associated SNP is rs9653442 (ref. 40)). The *IL2-IL21* SNP tested in our study, rs13119723, is in LD ($r^2 = 0.67$) with a SNP previously implicated in both Celiac disease and rheumatoid arthritis (rs6822844)[10,29] but is only in partial LD ($r^2 = 0.09$) with a T1D SNP (rs4505848)[23]. The *CD19-NFATC2IP* SNP tested in our study, rs8045689, was selected because of GRAIL analysis; it is in partial LD ($r^2 = 0.38$) with a SNP associated with T1D (rs4788084)[23]. The *TNIP1* SNP in our study, rs6889239, is in strong LD with an SLE SNP (rs7708392, $r^2 = 0.91$)[57] but is not in LD with another *TNIP1* SNP associated with psoriasis (rs17728338, $r^2 < 0.01$)[26]. The *ZEB1* SNP (rs2793108) was from the May 2009 release of T1D base, although this SNP did not appear in a subsequent publication[23]. The *PXK* SLE SNP was tested in this study but is not shown, as it is in weak LD with the rheumatoid arthritis risk SNP ($r^2 = 0.15$ between rs6445975 and rs13315591); the

© 2010 Nature America, Inc. All rights reserved.

rheumatoid arthritis SNP was selected because of $P_{\mathrm{GWAS}} < 10^{-6}$, not because of its association with another autoimmune disease.

49. Arnett, F.C. *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988).
50. Neale, B.M. *et al.* Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (LIPC). *Proc. Natl. Acad. Sci. USA* **107**, 7395–7400 (2010).
51. Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
52. Coenen, M.J. *et al.* Common and different genetic background for rheumatoid arthritis and coeliac disease. *Hum. Mol. Genet.* **18**, 4195–4203 (2009).
53. Wijbrandts, C.A. *et al.* The clinical response to infliximab in rheumatoid arthritis is in part dependent on pretreatment tumour necrosis factor alpha expression in the synovium. *Ann. Rheum. Dis.* **67**, 1139–1144 (2008).
54. Costenbader, K.H., Chang, S.C., De Vivo, I., Plenge, R. & Karlson, E.W. *PTPN22*, *PADI4* and *CTLA4* genetic polymorphisms and risk of rheumatoid arthritis in two longitudinal cohort studies: evidence of gene-environment interactions with heavy cigarette smoking. *Arthritis Res. Ther.* **10**, R52 (2008).
55. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
56. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
57. Gateva, V. *et al.* A large-scale replication study identifies *TNIP1*, *PRDM1*, *JAZF1*, *UHRF1BP1* and *IL10* as risk loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1228–1233 (2009).

© 2010 Nature America, Inc. All rights reserved.

doi:10.1038/ng.582