nature genetics

Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants

Wellcome Trust Case Control Consortium[1] & The Australo-Anglo-American Spondylitis Consortium[1]

We have genotyped 14,436 nonsynonymous SNPs (nsSNPs) and 897 major histocompatibility complex (MHC) tag SNPs from 1,000 independent cases of ankylosing spondylitis (AS), autoimmune thyroid disease (AITD), multiple sclerosis (MS) and breast cancer (BC). Comparing these data against a common control dataset derived from 1,500 randomly selected healthy British individuals, we report initial association and independent replication in a North American sample of two new loci related to ankylosing spondylitis, *ARTS1* and *IL23R*, and confirmation of the previously reported association of AITD with *TSHR* and *FCRL3*. These findings, enabled in part by increased statistical power resulting from the expansion of the control reference group to include individuals from the other disease groups, highlight notable new possibilities for autoimmune regulation and suggest that *IL23R* may be a common susceptibility factor for the major 'seronegative' diseases.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

Genome-wide association scans are currently revealing a number of new genetic variants for common diseases[1–11]. We have recently completed the largest and most comprehensive scan conducted to date, involving genome-wide association studies of 2,000 individuals from each of seven common disease cohorts and 3,000 common control individuals using a dense panel of >500,000 markers[12]. In parallel with this scan, we conducted a study of 5,500 independent individuals with a genome-wide set of nonsynonymous coding variants, an approach that has recently yielded new findings about type 1 diabetes and Crohn's disease and that has been proposed as an efficient complementary approach to whole-genome scans[13–15]. Here we report several new replicated associations in our scan of nsSNPs in 1,500 shared controls and 1,000 individuals from each of four different diseases: ankylosing spondylitis, AITD (of which all had Graves' disease), breast cancer and multiple sclerosis.

RESULTS

Initial genotyping was carried out with a custom-made Infinium array (Illumina) and involved 14,436 nsSNPs (assays were synthesized for 16,078 nsSNPs). At the inception of the study, this comprised the complete set of experimentally validated nsSNPs with minor allele frequency (MAF) > 1% in western European samples. In addition, because three of the diseases were of autoimmune etiology, we also typed a dense set of 897 SNPs throughout the MHC that, together with 348 nsSNPs in this region, provided comprehensive tag SNP coverage ($r^2 \geq 0.8$ with all SNPs in ref. 16). Finally, 103 SNPs were typed in pigmentation genes specifically designed to differentiate between population groups. Similar to those from previous studies, our data revealed that detailed assessment of initial data is critical to the process of association inference, as biases in genotype calling lead to inflation of false-positive rates[12,17]. This inflation is exaggerated in nsSNP data, because nsSNPs tend to have lower allele frequencies than otherwise anonymous genomic SNPs, and genotype calling is often most difficult for rare alleles. If only cursory filtering had been applied in the present case, numerous false-positives would have emerged (**Supplementary Figs. 1–4** online). **Table 1** shows the total number of SNPs and individuals remaining after genotype and sample quality control procedures (see Methods).

Association with the MHC

The strongest associations observed in the study were between SNPs in the MHC region and the three autoimmune diseases studied—ankylosing spondylitis, AITD and MS—with P values of $<10^{-20}$ for each disease (**Fig. 1**). No association of the MHC was seen with breast cancer ($P > 10^{-4}$ across the region). For each of the autoimmune diseases, the maximum signal was centered around the known HLA-associated genes (for example, those encoding HLA-B in ankylosing spondylitis, HLA-DRB1 in MS and the MHC class I and class II molecules in AITD), but in all cases, it extended far beyond the specific associated haplotype(s). For example, in ankylosing spondylitis, association was observed at $P < 10^{-20}$ across ~1.5 Mb. Given the well-known strong effect of HLA-B27 variant on the probability of developing ankylosing spondylitis (odds ratio 100–200 in most populations), the extent of this association signal reflects that with such large effects, even very distant SNPs in modest linkage disequilibrium (LD) will show indirect evidence for association. Strong signals like these may also cloud the evidence for additional HLA loci[18]. Disentangling similar patterns of association within the MHC has proven extremely challenging in the past and will be addressed in future studies of these data. Here we focus specifically on the nsSNP results.

[1]The complete lists of participants and affiliations appear at the end of the article. Correspondence should be addressed to L.R.C. (lcardon@fhcrc.org) or D.M.E. (davide@well.ox.ac.uk).

Received 17 July; accepted 17 September; published online 21 October 2007; doi:10.1038/ng.2007.17

Table 1 Number of individuals and SNPs tested in each cohort

	Cohort				
	AS	AITD	BC	MS	58C
Males	610	138	0	271	732
Females	312	762	1,004	704	734
Number of SNPs genotyped	15,436	15,436	15,436	15,436	15,436
SNPs with low GC score	783	816	771	802	796
SNPs with low genotyping	133	206	124	218	186
Monomorphic SNPs	1,842	1,829	1,854	1,810	1,687
SNPs with HW $P < 10^{-7}$[a]	129	74	104	97	132
Differences in missing rate $P < 10^{-4}$	51	101	172	309	n/a
'Manual' exclusions	33	33	33	33	33
Total number of SNPs tested	12,701	12,572	12,577	12,374	

[a]Only SNPs with HW $P < 10^{-7}$ in the 1958 birth cohort (58C) control group were excluded from analyses.

Association with nsSNPs

A major advantage of the Wellcome Trust Case Control Consortium (WTCCC) design is the availability of multiple disease cohorts that are similar in terms of ancestry and that have been typed on the same genetic markers[12,17]. Assuming that each disease has at least some unique genetic loci, we hypothesized that combining the other three case groups with the controls for the 1958 birth cohort (58C)[19] would increase power to detect association. For each disease, we therefore conducted two primary analyses: first, we tested nsSNP associations for each disease against the controls in the 58C; and second, we tested the same associations for each disease against an expanded reference group comprising the combined cases from the other three disease groups plus individuals from the 58C. A similar set of analyses was conducted for each of the autoimmune disorders against a reference group comprising 58C controls and individuals with breast cancer, but the results were very similar to those for the fully expanded groups, so here we describe the larger sample (**Supplementary Table 1** online). In addition, because it is possible that different autoimmune diseases share similar genetic etiologies, we also compared a combined ankylosing spondylitis, AITD and MS group (immune cases) against the combined set of individuals with breast cancer and 58C controls. All of our analyses are reported without regard to specific treatment of population structure, as the degree of structure in our final genotype data is not severe (Genomic Control[20] $\lambda = 1.07$–1.13 in the 58C-only datasets; $\lambda = 1.03$–1.06 in the expanded reference group comparisons; **Table 2**), consistent with our recent findings from 17,000 UK individuals involving the same controls[12].

nsSNP association results (excluding the MHC region) for each of the four disease groups against the 58C controls are shown in **Figure 2** and **Table 3**. Two SNPs on chromosome 5 reached a high level of statistical significance for ankylosing spondylitis (rs27044: $P = 1.0 \times 10^{-6}$; rs30187: $P = 3.0 \times 10^{-6}$). This level of significance exceeds the 10^{-5}–10^{-6} thresholds advocated for gene-based scans[21], as well as the oft-used Bonferroni correction at $P < 0.05$ (see refs. 12,21 for a discussion of genome-wide association significance). Both of these markers reside in the gene *ARTS1* (*ERAAP, ERAP1*), which encodes a type II integral transmembrane aminopeptidase with diverse immunological functions. Four additional SNPs show significance at $P < 10^{-4}$, with an increasing number of possible associations at more modest significance levels. Several of the more strongly associated SNPs, and others in the same genes, have been previously associated with these particular diseases, and for yet others there exists functional evidence of involvement in these particular conditions. Among these are SNPs in *FCRL3* and *FCRL5* in the case of AITD, *IL23R* in the case of ankylosing spondylitis, *MEL18* in the case of breast cancer and *IL7R* for MS. The complete list of single-marker association results is provided in **Supplementary Table 1**.

The results of analyses involving the expanded reference group are presented in **Supplementary Figure 5** online and **Supplementary**

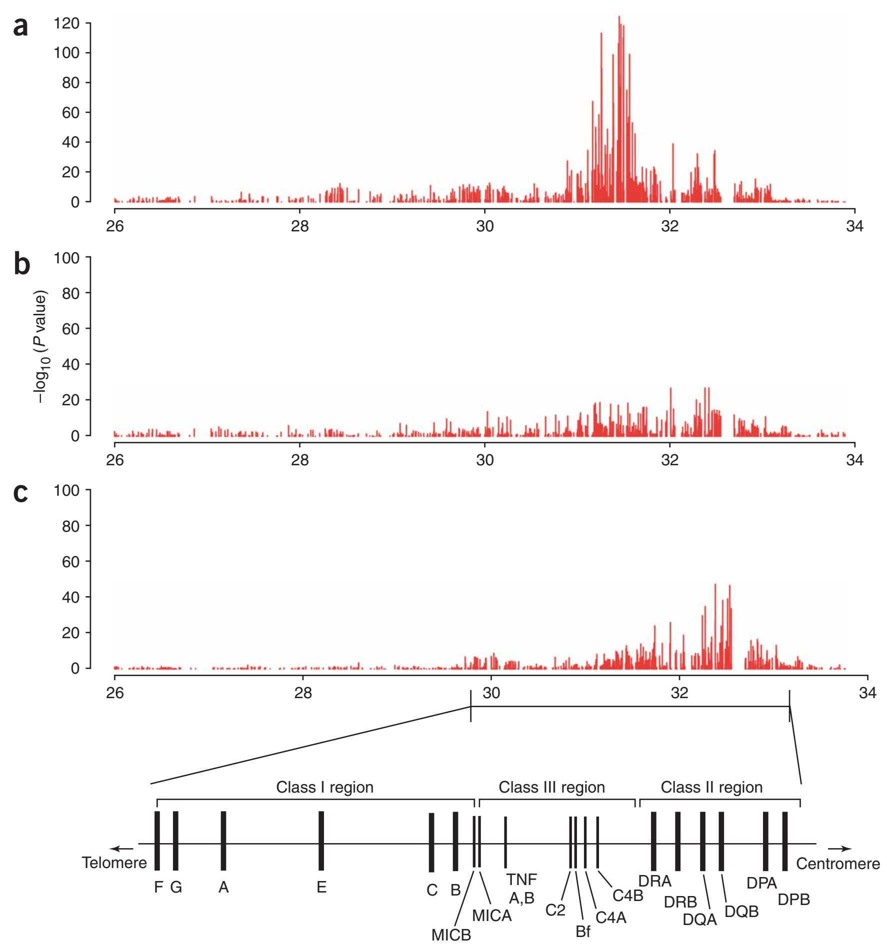

Figure 1 Minus $\log_{10}$ *P* values for the Armitage test of trend for MHC association with ankylosing spondylitis (**a**), autoimmune thyroid disease (**b**) and multiple sclerosis (**c**). Note in particular how evidence for association extends along very long regions of the MHC, reflecting statistical power to detect association even when linkage disequilibrium amongst SNPs is relatively low or when there exists the possibility of multiple disease-predisposing loci.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

Table 2 Estimates of λ for single and combined cohorts

		λ
Single cohort	AS cases versus 58C	1.07
	AITD cases versus 58C	1.12
	BC cases versus 58C	1.13
	MS cases versus 58C	1.12
Mixed cohorts	AS cases versus all others	1.03
	AITD cases versus all others	1.05
	BC cases versus all others	1.04
	MS cases versus all others	1.06
	IMMUNE cases versus BC and 58C	1.04

Table 1. Many of the SNPs that showed moderate to strong evidence for association in the initial analysis had substantially greater significance when the larger reference group was used. Notably, these included the SNPs rs27044 ($P = 4.0 \times 10^{-8}$) and rs30187 ($P = 2.1 \times 10^{-7}$) in *ARTS1*, as well as several other variants in this gene. A second SNP, rs7302230 in the gene encoding calsyntenin-3 on chromosome 12, showed substantially stronger evidence for association in the expanded reference group analysis ($P = 5.3 \times 10^{-7}$) relative to the 58C-only results ($P = 1.1 \times 10^{-4}$). Results of the expanded group also showed elevated results for several SNPs that did not appear exceptional in the original (non-combined) analyses, including SNPs in several candidate genes such as those encoding sialoadhesin[22] and complement receptor 1 for ankylosing spondylitis, *PIK3R2* for MS, and *C8B*, *IL17R* and *TYK2* in the combined autoimmune disease analysis. SNP rs3783941 in the gene *TSHR*, encoding the thyroid-stimulating hormone receptor, emerged as among the most significant in the expanded reference group analyses of AITD ($P = 2.1 \times 10^{-5}$). Several polymorphisms in *TSHR* have previously been associated with Graves' disease[23,24]. This known association did not reach even the modest significance level of 10^{-3} in the original analyses, but the addition of 3,000 further reference samples delineated it from the background noise and further supports the original independent report.

ARTS1 association confirmed in an independent cohort

To validate the most exceptional findings from the initial study, we genotyped the *ARTS1*, *CLSTN3* and *LNPEP* SNPs in 471 independent ankylosing spondylitis cases (**Table 4**) and 625 new controls (all self identified North American Caucasian). The data strongly suggest that the *ARTS1* association is genuine. All *ARTS1* nsSNPs revealed independent replication in the same direction of effect, with replication significance levels ranging from 4.7×10^{-4} to 5.1×10^{-5}. When combined with the original samples, the results showed strong evidence for association with ankylosing spondylitis ($P = 1.2 \times 10^{-8}$ to 3.4×10^{-10}). The population attributable risk[25] contributed by the most strongly associated marker in the North American dataset (rs2287987) was 26%.

Association was also confirmed with marker rs2303138 in the *LNPEP* gene, which lies 127 kb 3′ of *ARTS1*. This marker was in strong LD with *ARTS1* markers ($D' = 1$, rs27044–rs2303138). We tested the interdependence of the *ARTS1* and *LNPEP* associations using conditional logistic regression. The remaining association at *LNPEP* was weak after controlling for *ARTS1* ($P = 0.01$), whereas the association at *ARTS1* remained strong after controlling for *LNPEP* ($P = 2.7 \times 10^{-6}$), suggesting that the *LNPEP* association may only be secondary to LD, with a true association at *ARTS1*.

No association was seen with *CLSTN3* in the confirmation set. The US controls showed the same allele frequency as the UK controls (5%), but the allele frequency in the US cases was less than that of the UK cases (6% versus 8%), suggesting no association in the US samples and substantially reducing the significance of the combined data. Calystenin-3 is a postsynaptic neuronal membrane protein and is an unlikely candidate for involvement in inflammatory arthritis. The failure to replicate this association suggests that our replication sample size was insufficient to detect the modest effect or that it was a false positive in the initial scan.

IL23R variants confer risk of ankylosing spondylitis

The *IL23R* variant rs11209026, although not notable in the initial nsSNP scan ($P = 1.7 \times 10^{-3}$), was of particular interest, as it has recently been associated with both Crohn's disease[26,27] and psoriasis[28], conditions that commonly co-occur with ankylosing spondylitis. To better define this association, seven additional SNPs in *IL23R* were genotyped in the same 1,000 British ankylosing spondylitis cases and 1,500 58C controls as well as the North American Caucasian replication samples (**Table 4**). In the WTCCC dataset, we observed strong association in seven of eight genotyped SNPs ($P \leq 0.008$, including the original nsSNP rs11209026), with the strongest association at rs11209032 ($P = 2.0 \times 10^{-6}$). In the replication dataset, we noted association with all genotyped SNPs ($P \leq 0.04$), with peak association with marker rs10489629 ($P = 4.2 \times 10^{-5}$). In the combined dataset,

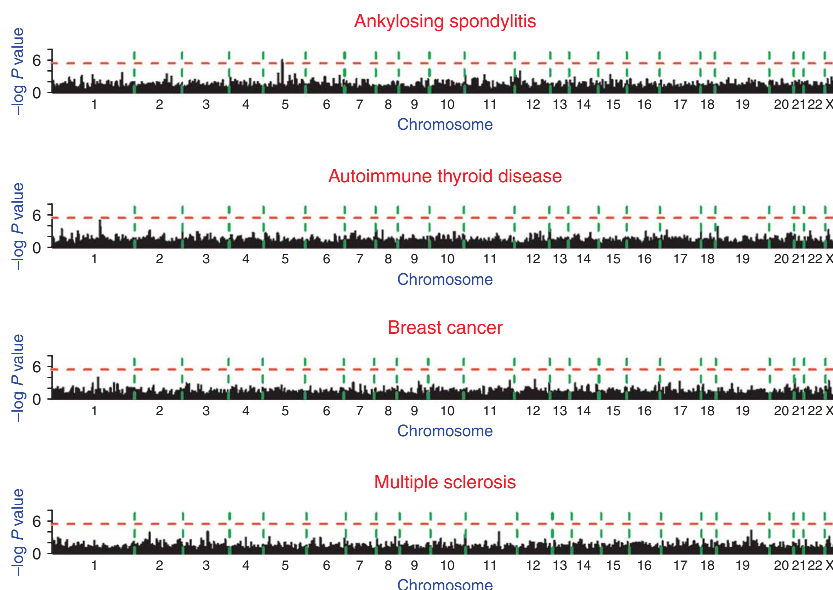

Figure 2 Minus $\log_{10}$ *P* values for the Armitage test of trend for genome-wide association scans of ankylosing spondylitis, autoimmune thyroid disease, breast cancer and multiple sclerosis. The spacing between SNPs on the plot is uniform and does not reflect distances between the SNPs. The vertical dashed lines reflect chromosomal boundaries. The horizontal dashed lines display the cutoff of $P = 10^{-6}$. Note that SNPs within the MHC are not included in this diagram.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

the strongest association observed was with SNP rs11209032 (odds ratio 1.3, 95% confidence interval 1.2–1.4, $P = 7.5 \times 10^{-9}$). The attributable risk for this marker in the replication cohort is 9%. Conditional logistic regression analyses did not indicate a single primary disease-associated marker; residual association remained after we controlled for association at the remaining SNPs. Considering only individuals with ankylosing spondylitis who self-reported as not having inflammatory bowel disease ($n = 1{,}066$) the association remained strong and was still strongest at rs11209032 ($P = 6.9 \times 10^{-7}$), indicating that there is a primary association with ankylosing spondylitis and that the observed association was not due to coexistent clinical inflammatory bowel disease.

In contrast to the pleiotropic effects of *IL23R*, the *ARTS1* association evidence seems confined to ankylosing spondylitis. We genotyped

Table 3 nsSNPs outside the MHC that meet a point-wise significance level of $P < 10^{-3}$ for the Cochran-Armitage test for trend

Disease	SNP	Chromosome	Position (bp)	MAF	OR	χ^2	*P* value	Gene
AS	rs696698	1	74777462	0.04	1.84	11.13	8.5×10^{-4}	*C1orf173*
	rs10494217	1	119181230	0.17	0.77	11.62	6.5×10^{-4}	*TBX15*
	rs2294851	1	206966279	0.13	0.73	13.55	2.3×10^{-4}	*HHAT*
	rs8192556	2	182368504	0.01	0.45	12.24	4.7×10^{-4}	*NEUROD1*
	rs16876657	5	78645930	0.02	3.10	13.05	3.0×10^{-4}	*JMY*
	rs27044	5	96144608	0.34	1.40	23.90	1.0×10^{-6}	*ARTS-1*
	rs17482078	5	96144622	0.17	0.76	13.55	2.3×10^{-4}	*ARTS-1*
	rs10050860	5	96147966	0.18	0.75	14.87	1.1×10^{-4}	*ARTS-1*
	rs30187	5	96150086	0.40	1.33	21.82	3.0×10^{-6}	*ARTS-1*
	rs2287987	5	96155291	0.18	0.75	14.31	1.6×10^{-4}	*ARTS-1*
	rs2303138	5	96376466	0.10	1.58	19.41	1.1×10^{-5}	*LNPEP*
	rs11750814	5	137528564	0.16	0.77	10.99	9.1×10^{-4}	*BRD8*
	rs11959820	5	149192703	0.02	0.49	12.41	4.3×10^{-4}	*PPARGC1B*
	rs907609	11	1813846	0.13	0.76	10.91	9.5×10^{-4}	*SYT8*
	rs3740691	11	47144987	0.29	0.80	11.86	5.7×10^{-4}	*ZNF289*
	rs11062385	12	297836	0.24	0.79	11.82	5.9×10^{-4}	*JARID1A*
	rs7302230	12	7179699	0.08	1.57	14.97	1.1×10^{-4}	*CLSTN3*
AITD	rs10916769	1	20408244	0.17	0.76	12.10	5.0×10^{-4}	*FLJ32784*
	rs6427384	1	154321955	0.18	1.43	18.97	1.3×10^{-5}	*FCRL5*
	rs2012199	1	154322098	0.17	1.35	13.18	2.8×10^{-4}	*FCRL5*
	rs6679793	1	154327170	0.22	1.33	14.69	1.3×10^{-4}	*FCRL5*
	rs7522061	1	154481463	0.47	1.25	13.78	2.1×10^{-4}	*FCRL3*
	rs1047911	2	74611433	0.15	1.34	11.24	8.0×10^{-4}	*MRPL53*
	rs7578199	2	241912838	0.26	1.26	11.53	6.9×10^{-4}	*HDLBP*
	rs3748140	8	9036429	0.00	0.28	11.44	7.2×10^{-4}	*PPP1R3B*
	rs1048101	8	26683945	0.42	0.82	10.98	9.2×10^{-4}	*ADRA1A*
	rs7975069	12	132389146	0.30	0.80	12.06	5.2×10^{-4}	*ZNF268*
	rs2271233	17	6644845	0.07	0.94	11.32	7.7×10^{-4}	*TEKT1*
	rs2856966	18	897710	0.19	0.76	14.00	1.8×10^{-4}	*ADCYAP1*
	rs7250822	19	2206311	0.04	1.97	13.83	2.0×10^{-4}	*AMH*
	rs2230018	23	44685331	0.14	1.41	11.55	6.8×10^{-4}	*UTX*
BC	rs4255378	1	151919300	0.48	1.25	14.70	1.3×10^{-4}	*MUC1*
	rs2107732	7	44851218	0.10	1.40	10.96	9.3×10^{-4}	*CCM2*
	rs4986790	9	117554856	0.07	1.54	11.46	7.1×10^{-4}	*TLR4*
	rs2285374	11	118457383	0.38	0.82	12.25	4.7×10^{-4}	*VPS11*
	rs7313899	12	54231386	0.03	2.10	13.02	3.1×10^{-4}	*OR6C4*
	rs2879097	17	34143085	0.20	0.78	11.73	6.1×10^{-4}	*MEL18*
	rs2822558	21	14593715	0.13	0.73	13.87	2.0×10^{-4}	*ABCC13*
	rs2230018	23	44685331	0.14	1.40	12.14	4.9×10^{-4}	*UTX*
MS	rs17009792	2	74400978	0.02	0.44	14.41	1.5×10^{-4}	*SLC4A5*
	rs1132200	3	120633526	0.15	0.73	15.22	9.6×10^{-5}	*FLJ10902*
	rs6897932	5	35910332	0.23	0.80	11.04	8.9×10^{-4}	*IL7R*
	rs6470147	8	124517985	0.36	1.23	10.92	9.5×10^{-4}	*FLJ10204*
	rs3818511	10	134309378	0.24	1.28	12.84	3.4×10^{-4}	*INPP5A*
	rs11574422	11	67970565	0.02	2.82	14.64	1.3×10^{-4}	*LRP5*
	rs388706	19	49110533	0.48	1.22	11.19	8.2×10^{-4}	*ZNF45*
	rs1800437	19	50873232	0.17	0.74	16.11	6.0×10^{-5}	*GIPR*
	rs2281868	23	69451484	0.50	1.26	11.38	7.4×10^{-4}	*SAP102*

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

Table 4 Ankylosing spondylitis replication results

		UK cases				US cases				All cases			
Gene	SNP	Case MAF	Control MAF	OR	*P* value	Case MAF	Control MAF	OR	*P* value	Case MAF	Control MAF	OR	*P* value
ARTS1	rs27044	0.34	0.27	1.40	1.0×10^{-6}	–	–	–	–	–	–	–	–
ARTS1	rs17482078	0.17	0.22	0.76	2.3×10^{-4}	0.15	0.21	0.65	5.1×10^{-5}	0.16	0.22	0.70	1.2×10^{-8}
ARTS1	rs10050860	0.18	0.23	0.75	1.2×10^{-4}	0.15	0.22	0.66	8.8×10^{-5}	0.17	0.22	0.71	7.6×10^{-9}
ARTS1	rs30187	0.40	0.33	1.33	3.0×10^{-6}	0.41	0.35	1.30	0.00047	0.41	0.34	1.40	3.4×10^{-10}
ARTS1	rs2287987	0.18	0.22	0.75	1.6×10^{-4}	0.15	0.21	0.66	8.4×10^{-5}	0.17	0.22	0.71	1.0×10^{-8}
LNPEP	rs2303138	0.10	0.07	1.58	1.1×10^{-5}	0.11	0.09	1.40	0.018	0.11	0.07	1.48	1.1×10^{-6}
CLSTN3	rs7302230	0.08	0.05	1.57	1.1×10^{-4}	0.06	0.05	1.10	0.56	0.07	0.05	1.30	0.0039
IL23R	rs11209026	0.04	0.06	0.63	0.0017	0.038	0.06	0.63	0.014	0.04	0.06	0.63	4.0×10^{-6}
IL23R	rs1004819	0.35	0.30	1.20	0.0013	0.35	0.30	1.30	0.0045	0.35	0.30	1.20	1.1×10^{-5}
IL23R	rs10489629	0.43	0.45	0.90	0.062	0.39	0.47	0.72	4.2×10^{-5}	0.41	0.46	0.83	0.00011
IL23R	rs11465804	0.04	0.06	0.67	0.0019	0.049	0.06	0.68	0.04	0.04	0.06	0.68	0.0002
IL23R	rs1343151	0.30	0.34	0.85	0.0077	0.29	0.36	0.71	6.7×10^{-5}	0.30	0.34	0.80	1.0×10^{-5}
IL23R	rs10889677	0.36	0.31	1.20	0.00066	0.37	0.29	1.40	4.7×10^{-5}	0.36	0.31	1.30	1.3×10^{-6}
IL23R	rs11209032	0.38	0.32	1.30	2.0×10^{-6}	0.38	0.32	1.30	0.0013	0.38	0.32	1.30	7.5×10^{-9}
IL23R	rs1495965	0.49	0.44	1.20	0.0021	0.50	0.43	1.40	0.00019	0.49	0.44	1.20	3.1×10^{-6}

the five ankylosing spondylitis–associated SNPs in 755 British Crohn's disease and 1,011 ulcerative colitis cases and 633 healthy controls. No association was seen with either ulcerative colitis or Crohn's disease (Armitage trend $P > 0.4$ for all markers).

FCRL3 confirmed in AITD pathogenesis

In addition to the ankylosing spondylitis replications, we attempted to confirm and extend the *FCRL3* association in AITD. The SNP rs7522061 in the *FCRL3* gene was recently reported to be associated with AITD[29] and two other autoimmune diseases, rheumatoid arthritis and systemic lupus erythematosus[30]. Our initial association evidence ($P = 2.1 \times 10^{-4}$) likely reflects the signal of the originally detected polymorphism, because the level of LD is high across this gene. In fact, the entire 1q21–q23 region (which includes another gene, *FCRL5*, flagged in our scan) has also been implicated in several autoimmune diseases, including psoriasis and multiple sclerosis[31,32].

On the basis of the original findings on 1q21–q23, the original cohort was increased from 1,000 to 2,500 Graves disease cases, and we used 2,500 controls from the 58C control set. We selected eight SNPs that tagged the *FCRL3* and *FCRL5* gene regions and typed them in all 5,000 samples using an alternative genotyping platform. SNP rs3761959, which tags rs7522061 and rs7528684 (previously associated with rheumatoid arthritis and Graves' disease), was associated with Graves' disease in this extended cohort (**Table 5**), confirming the original result. In total, three of the seven *FCRL3* SNPs showed some evidence for association ($P < 0.05$), with SNP rs11264798 showing the strongest association of the tag SNPs ($P = 4.0 \times 10^{-3}$). SNP rs6667109 in *FCRL5*, which tagged SNPs rs6427384, rs2012199 and rs6679793, all found to be weakly associated in the original study, showed little evidence of association in this extended cohort.

DISCUSSION

Our scan of nsSNPs has identified and validated two new genes (*ARTS1* and *IL23R*) associated with ankylosing spondylitis, confirmed and extended markers in the *TSHR* and *FCRL3* genes that have previously been associated with AITD, and provided a dense set of association data for AITD, ankylosing spondylitis and MS across the MHC region. The challenge now is to design functional studies that

Table 5 Autoimmune thyroid disease replication results

		Replication cohort				Combined cohort			
Gene	SNP	Case MAF	Control MAF	OR	*P* value	Case MAF	Control MAF	OR	*P* value
FCRL3	rs3761959[a]	0.48	0.45	0.87	0.013	0.49	0.45	0.87	9.4×10^{-3}
FCRL3	rs11264794	0.42	0.45	1.10	0.079	0.42	0.46	1.12	0.013
FCRL3	rs11264793	0.27	0.24	0.87	0.029	0.26	0.24	0.90	0.044
FCRL3	rs11264798	0.44	0.49	1.18	4.0×10^{-3}	0.44	0.49	1.22	1.6×10^{-5}
FCRL3	rs10489678	0.19	0.20	1.04	0.58	0.20	0.20	1.04	0.43
FCRL3	rs6691569	0.28	0.29	1.02	0.75	0.29	0.29	1.00	0.93
FCRL3	rs2282284	0.062	0.058	0.92	0.015	0.062	0.058	0.93	0.47
FCRL5	rs6667109	0.17	0.16	0.93	0.38	0.18	0.15	0.85	7.7×10^{-2}

[a]This SNP tags the SNP rs7522061, which was flagged as associated with AITD in the WTCCC screen ($P = 2.1 \times 10^{-4}$).

will reveal how variation in these genes translates into physiological processes that influence disease risk.

From a functional perspective, *ARTS1* and *IL23R* represent excellent biological candidates for association with ankylosing spondylitis. The protein ARTS1 has two known functions, either of which may explain the association. First, within the endoplasmic reticulum, ARTS1 is involved in trimming peptides to the optimal length for MHC class I presentation[33,34]. Ankylosing spondylitis is primarily an HLA class I–mediated autoimmune disease[35], with >90% of cases carrying the HLA-B27 allele. How HLA-B27 increases risk of developing ankylosing spondylitis is unknown, but if the association of *ARTS1* with the disease relates to effects of ARTS1 on peptide presentation, this relationship would inform research into the mechanism underlying the association of HLA-B27 with ankylosing spondylitis. Second, ARTS1 cleaves cell surface receptors for the pro-inflammatory cytokines IL-1 (IL-1R2)[36], IL-6 (IL-6Rα)[37] and TNF (TNFR1)[38], thereby downregulating their signaling. Genetic variants that alter the functioning of ARTS1 could therefore have pro-inflammatory effects through this mechanism.

In addition to their association with ankylosing spondylitis, polymorphisms in *IL23R* have been recently documented in Crohn's disease[26,27] and psoriasis[28], suggesting that this gene is a common susceptibility factor for the major 'seronegative' diseases, at least partially explaining their co-occurrence. IL-23R is a key factor in the regulation of a newly defined effector T-cell subset, T_H17 cells. T_H17 cells were originally identified as a distinct subset of T-cells expressing high levels of the pro-inflammatory cytokine IL-17 in response to stimulation, in addition to IL-1, IL-6, TNFα, IL-22 and IL-25 (IL-17E). IL-23 has been shown to be important in the mouse models of experimental autoimmune encephalomyelitis[39], collagen-induced arthritis[40] and inflammatory bowel disease[41], but it has not been studied in ankylosing spondylitis, either in human or other animal models of the disease. These studies show that blocking IL-23 reduces inflammation in these models, suggesting that the *IL23R* variants associated with disease are pro-inflammatory. Successful treatment of Crohn's disease has been reported with anti–IL-12p40 antibodies, which block both IL-12 and IL-23, as these cytokines share the IL-12p40 chain[42]. No functional studies of *IL23R* variants have been reported to date, and it is unclear to what extent findings in studies targeting IL-23 can be generalized to mechanisms by which *IL23R* variation affects disease susceptibility. Our genetic findings provide notable insight into the etiopathogenesis of ankylosing spondylitis and suggest that treatments targeting IL-23 may prove effective for this condition, but clearly much more needs to be understood about the mechanism underlying the observed association.

Despite the successful identification of the *ARTS1* and *IL23R* genes, it is likely either that additional real associations are present in our data but were overlooked because of their modest effect sizes, or that our focus on non-synonymous coding changes led us to miss real loci. The issue of limited statistical power is emphasized in studies of nonsynonymous coding changes, which have a greater number of rare variants than other genetic variants and thus will require even larger sample sizes unless the effect sizes are larger. Other analytical approaches, such as assessing evidence for association between clusters of rare variants rather than individual loci, may prove highly informative in this regard[43], but most of the nsSNPs available in this study exist either by themselves in each gene or with one or two others, which precludes these assessments (**Supplementary Fig. 6** online). In our analyses, *ARTS1* was the only locus showing exceptional statistical significance in the scan of 1,000 cases and 1,500 controls, thus emphasizing the need for greater statistical power. We increased power by expanding the controls, or 'reference set,' to include some or all of the other disease samples. When we did so, *ARTS1* showed even stronger association evidence, the *IL23R* SNPs increased to a level that began to delineate them from background noise, and the AITD/*TSHR* confirmation emerged. This demonstration of increased statistical power through the combination of multiple datasets is timely, given the international impetus to make genotype data available to the scientific community. Future investigations will be needed to assess the power versus confounding effects and the statistical corrections needed to combine more heterogeneous samples from broader sampling regions.

These results also highlight the question of how much information may be missed by focusing on coding SNPs rather than searching more broadly over the genome at large. This question is relevant because the tradeoff between SNP panel and sample size selection is a salient factor in the design of every genome-wide study. In the HapMap data[44], a substantial portion of the common nonsynonymous variation in our nsSNP set is captured by available genome-wide panels (about 65% of common (MAF > 5%) nsSNPs in the Illumina Human NS-12 Beadchip are tagged with an $r^2 > 0.8$ using the Affymetrix 500 K chip, rising to 90% in the Illumina HumanHap300, which includes almost all of the nsSNPs from the NS-12 Beadchip). The four primary associated variants flagged in our study (that is, in *ARTS1*, *IL23R*, *TSHR* and *FCRL3*) would have been detected using any of the genome-wide panels, because either the markers themselves or a SNP in high LD with them ($r^2 \geq 0.78$) are present on the genome-wide chips. This LD relationship also emphasizes the fact that observing an association with a nsSNP does not necessarily imply that the nsSNP is causal, as it may be indirectly associated with other genetic variants in or outside the gene. Given this high degree of overlap, the continuously increasing coverage of many available genotyping products and concomitant pressures to decrease assay costs, these data suggest that future gene-centric scans will be efficiently subsumed by the more encompassing and less hypothesis-driven genome-wide SNP panels.

METHODS

Subjects. Individuals included in the study were self-identified as white and of European ancestry and came from mainland UK (England, Scotland and Wales, but not Northern Ireland). The 1,500 control samples were from the British 1958 Birth Cohort (58C, also known as the National Child Development Study), which included all the births in England, Wales and Scotland that occurred during 1 week in 1958. Recruitment details and diagnostic criteria for each of the four case groups, as well as for the North American AS replication cohort and the 58C are further described in the **Supplementary Methods** online.

Sample quality assurance and control genome-wide identity by state (IBS) sharing was calculated for each pair of individuals in the combined sample of cohorts to identify first- and second-degree relatives whose data might contaminate the study. One subject from any pair of individuals who shared <400 genotypes IBS = 0 and/or >80% alleles IBS (that is, the individual with the most missing genotypes) was removed from all subsequent analyses. To identify individuals who might have ancestries other than Western European, we merged each of our cohorts with the 60 western European (CEU) founder, 60 Nigerian (YRI) founder, and 90 Japanese (JPT) and Han Chinese (CHB) individuals from the International HapMap Project[44]. We calculated genome-wide IBD distances for each pair of individuals (that is, 1 minus average IBS sharing) on those markers shared between HapMap and our nonsynonymous panel, and then used the multidimensional scaling option in R to generate a two dimensional plot based upon individuals' scores on the first two principal coordinates from this analysis (**Supplementary Fig. 2**). Any WTCCC sample that was not present in the main cluster with the CEU individuals was excluded from subsequent analyses. Finally, any individual with >10%

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

of genotypes missing was removed from the analysis. The number of individuals remaining after these quality control measures were applied is shown in **Table 1**.

Genotyping. We genotyped a total of 14,436 nsSNPs across the genome on all case and control samples. Because three of the diseases were of autoimmune etiology, we also typed an additional 897 SNPs within the MHC region, as well as 103 SNPs in pigmentation genes specifically designed to differentiate between population groups. SNP genotyping was performed with the Infinium I assay (Illumina), which is based on allele-specific primer extension (ASPE) and the use of a single fluorochrome. The assay requires $\sim$250 ng of genomic DNA, which is first subjected to a round of isothermal amplification that generates a 'high-complexity' representation of the genome with most loci represented at usable amounts. There are two allele-specific probes (50-mers) per SNP, each on a different bead type; each bead type is present on the array an average of 30 times (and a minimum of 5 times), allowing for multiple independent measurements. We processed six samples per array. Clustering was carried out with the GenCall software version 6.2.0.4, which assigns a quality score to each locus and an individual genotype confidence score that is based on the distance of a genotype from the center of the nearest cluster. First, we removed samples with more than 50% of loci having a quality score below 0.7 and then all loci with a quality score below 0.2. After clustering, we applied two additional filtering criteria: (i) we omitted individual genotypes with a genotype confidence score $<$0.15 and (ii) we removed any SNP for which more than 20% of samples had genotype confidence scores $<$0.15. The above criteria were designed to optimize genotype accuracy and minimize uncalled genotypes.

Statistical analysis markers that were monomorphic in both case and control samples, SNPs with $>$10% missing genotypes and SNPs with differences in the amount of missing data between cases and controls ($P < 10^{-4}$ as assessed by χ^2 test) were excluded from all analyses involving that case group only. In addition, any marker that failed an exact test of Hardy-Weinberg equilibrium in controls ($P < 10^{-7}$) was excluded from all analyses[45].

Cochran-Armitage tests for trend[46] were conducted using the PLINK program[47]. For the present analyses, we used the significance thresholds of $P < 10^{-4}$–10^{-6}, as suggested for gene-based scans with stronger prior probabilities than scans of anonymous markers[21]. In the present context, the lower thresholds are similar to Bonferroni significance levels (Bonferroni-corrected $P = 0.05$ corresponds to nominal $P = 3 \times 10^{-6}$). The conditional logistic regression analyses involving the *LNPEP* and *ARTS1* SNPs were carried out using Purcell's WHAP program[48].

We manually rechecked the genotype calls of every nsSNP with an asymptotic significance level of $P < 10^{-3}$ by inspecting raw signal intensity values and their corresponding automated genotype calls. Notably, this flagged an additional 33 markers with clear problems in genotype calling, which were subsequently excluded from all analyses (**Supplementary Fig. 4**). These results indicate that this genotyping platform generally yields highly accurate genotypes, but errors do occur and can be distributed nonrandomly between cases and controls despite stringent quality control procedures. It is imperative to check the clustering of the most significant SNPs to ensure that evidence for associations is not a result of genotyping error.

Although great lengths were taken to ensure that our samples were as homogenous as possible in terms of genetic ancestry, even subtle population substructure can substantially influence tests of association in large genome-wide analyses involving thousands of individuals[49]. We therefore calculated the genomic-control inflation factor, λ (ref. 20), for each case-control sample as well as in the analyses where we combined the other case groups with the control individuals (**Table 2**). In general, values for λ were small ($\sim$1.1), indicating a small degree of substructure in UK samples that induces only a slight inflation of the test statistic under the null hypothesis, consistent with the results from our companion paper[12]. We therefore present uncorrected results in all analyses reported.

Consent was granted from ethical review boards of the institutions with which the participants were affiliated, and informed consent was obtained from the individuals involved in the WTCCC. Individual-level data from this study will be widely available through the Consortium's Data Access Committee (http://www.wtccc.org.uk).

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

ACKNOWLEDGMENTS

We would like to thank the Wellcome Trust for supporting this study, and all the individuals and controls who participated in this study.

AITD: We thank the collection coordinators, J. Carr-Smith and all contributors to the AITD national DNA collection of index cases and family members from centres including Birmingham, Bournemouth, Cambridge, Cardiff, Exeter, Leeds, Newcastle and Sheffield. Principal leads for the AITD UK national collection are S.C. Gough (Birmingham), S.H.S. Pearce (Newcastle), B. Vaidya (Exeter), J.H. Lazarus (Cardiff), A. Allahabadia (Sheffield), M. Armitage (Bournemouth), P.J. Grant (Leeds) and V.K. Chatterjee (Cambridge).

Ankylosing spondylitis: We thank the Arthritis Research Campaign (UK). MAB is funded by the National Health and Medical Research Council (Australia). TASC is funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases grants 1PO1-052915-01,RO1 AR046208 and RO1-AR048465, as well as by University of Texas at Houston CTSA grant UL1RR024148, Cedars-Sinai GCRC grant MO1-RR00425, The Rosalind Russell Center for Arthritis Research at The University of California San Francisco, and the Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, US National Institutes of Health. We thank R. Jin for technical assistance and L. Diekman, L. Guthrie, F. Lin and S. Morgan for their study coordination.

Breast cancer: The breast cancer samples were clinically and molecularly curated with the assistance of A. Renwick, A. Hall, A. Elliot, H. Jayatilake, T. Chagtai, R. Barfoot, P. Kelly and K. Spanova. Our research is supported by United States Army Medical Research and Material Command grant no. W81XWH-05-1-0204, The Institute of Cancer Research and Cancer Research UK.

MS: Our work has been supported by the Wellcome Trust (grant ref. 057097), the Medical Research Council (UK) (grant ref. G0000648), the Multiple Sclerosis Society of Great Britain and Northern Ireland (grant ref. 730/02) and the National Institutes of Health (USA) (grant ref. 049477). A.G. is a postdoctoral fellow of the Research Foundation–Flanders (FWO–Vlaanderen).

L. Galver and P. Ng at Illumina and J. Morrison at the Sanger Institute contributed in the design of the nsSNP array. We thank the DNA team of the JDRF/WT DIL and T. Dibling, C. Hind and D. Simpkin at the Sanger Institute for carrying out the genotyping. We also thank S. Bingham and the WTCCC Inflammatory Bowel Disease group for genotyping the *ARTS1* markers in their replication samples.

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

1. Rioux, J.D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* **39**, 596–604 (2007).
2. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
3. Easton, D.F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
4. Libioulle, C. *et al.* Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.* **3**, e58 (2007).
5. Zanke, B.W. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* **39**, 989–994 (2007).
6. Haiman, C.A. *et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat. Genet.* **39**, 638–644 (2007).
7. Gudmundsson, J. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.* **39**, 977–983 (2007).
8. Moffatt, M.F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473 (2007).
9. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
10. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
11. Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
12. WTCCC. Genome-wide association studies of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–683 (2007).
13. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.* **39**, 207–211 (2007).
14. Jorgenson, E. & Witte, J.S. Coverage and power in genomewide association studies. *Am. J. Hum. Genet.* **78**, 884–888 (2006).
15. Smyth, D.J. *et al.* A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat. Genet.* **38**, 617–619 (2006).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

16. Miretti, M.M. *et al.* A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **76**, 634–646 (2005).
17. Clayton, D.G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* **37**, 1243–1246 (2005).
18. Sims, A.M. *et al.* Non-B27 MHC associations of ankylosing spondylitis. *Genes Immun.* **8**, 115–123 (2007).
19. McGinnis, R., Shifman, S. & Darvasi, A. Power and efficiency of the TDT and case-control design for association scans. *Behav. Genet.* **32**, 135–144 (2002).
20. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
21. Thomas, D.C. & Clayton, D.G. Betting odds and genetic associations. *J. Natl. Cancer Inst.* **96**, 421–423 (2004).
22. Jiang, H.R. *et al.* Sialoadhesin promotes the inflammatory response in experimental autoimmune uveoretinitis. *J. Immunol.* **177**, 2258–2264 (2006).
23. Dechairo, B.M. *et al.* Association of the TSHR gene with Graves' disease: the first disease specific locus. *Eur. J. Hum. Genet.* **13**, 1223–1230 (2005).
24. Hiratani, H. *et al.* Multiple SNPs in intron 7 of thyrotropin receptor are associated with Graves' disease. *J. Clin. Endocrinol. Metab.* **90**, 2898–2903 (2005).
25. Miettinen, O.S. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am. J. Epidemiol.* **99**, 325–332 (1974).
26. Duerr, R.H. *et al.* A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. *Science* (2006).
27. Tremelling, M. *et al.* IL23R variation determines susceptibility but not disease phenotype in inflammatory bowel disease. *Gastroenterology* **132**, 1657–1664 (2007).
28. Cargill, M. *et al.* A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *Am. J. Hum. Genet.* **80**, 273–290 (2007).
29. Simmonds, M.J. *et al.* Contribution of single nucleotide polymorphisms within FCRL3 and MAP3K7IP2 to the pathogenesis of Graves' disease. *J. Clin. Endocrinol. Metab.* **91**, 1056–1061 (2006).
30. 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).
31. Capon, F. *et al.* Fine mapping of the PSORS4 psoriasis susceptibility region on chromosome 1q21. *J. Invest. Dermatol.* **116**, 728–730 (2001).
32. Dai, K.Z. *et al.* The T cell regulator gene SH2D2A contributes to the genetic susceptibility of multiple sclerosis. *Genes Immun.* **2**, 263–268 (2001).
33. Chang, S.C., Momburg, F., Bhutani, N. & Goldberg, A.L. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. *Proc. Natl. Acad. Sci. USA* **102**, 17107–17112 (2005).
34. Saveanu, L. *et al.* Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat. Immunol.* **6**, 689–697 (2005).
35. Brown, M.A. *et al.* HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom. *Ann. Rheum. Dis.* **55**, 268–270 (1996).
36. Cui, X., Rouhani, F.N., Hawari, F. & Levine, S.J. Shedding of the type II IL-1 decoy receptor requires a multifunctional aminopeptidase, aminopeptidase regulator of TNF receptor type 1 shedding. *J. Immunol.* **171**, 6814–6819 (2003).
37. Cui, X., Rouhani, F.N., Hawari, F. & Levine, S.J. An aminopeptidase, ARTS-1, is required for interleukin-6 receptor shedding. *J. Biol. Chem.* **278**, 28677–28685 (2003).
38. Cui, X. *et al.* Identification of ARTS-1 as a novel TNFR1-binding protein that promotes TNFR1 ectodomain shedding. *J. Clin. Invest.* **110**, 515–526 (2002).
39. Cua, D.J. *et al.* Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. *Nature* **421**, 744–748 (2003).
40. Murphy, C.A. *et al.* Divergent pro- and antiinflammatory roles for IL-23 and IL-12 in joint autoimmune inflammation. *J. Exp. Med.* **198**, 1951–1957 (2003).
41. Hue, S. *et al.* Interleukin-23 drives innate and T cell-mediated intestinal inflammation. *J. Exp. Med.* **203**, 2473–2483 (2006).
42. Mannon, P.J. *et al.* Anti-interleukin-12 antibody for active Crohn's disease. *N. Engl. J. Med.* **351**, 2069–2079 (2004).
43. Cohen, J.C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869–872 (2004).
44. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
45. Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. A note on exact tests of Hardy-Weinberg equilibrium. *Am. J. Hum. Genet.* **76**, 887–893 (2005).
46. Armitage, P. Test for linear trend in proportions and frequencies. *Biometrics* **11**, 375–386 (1955).
47. 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).
48. Purcell, S., Daly, M.J. & Sham, P.C. WHAP: haplotype-based association analysis. *Bioinformatics* **23**, 255–256 (2007).
49. Marchini, J., Cardon, L.R., Phillips, M.S. & Donnelly, P. The effects of human population structure on large genetic association studies. *Nat. Genet.* **36**, 512–517 (2004).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

The complete list of authors is as follows:

Wellcome Trust Case Control Consortium

Management Committee: Paul R Burton[1], David G Clayton[2], Lon R Cardon[3,55], Nick Craddock[4], Panos Deloukas[5], Audrey Duncanson[6], Dominic P Kwiatkowski[3,5], Mark I McCarthy[3,7], Willem H Ouwehand[8,9], Nilesh J Samani[10], John A Todd[2], Peter Donnelly (Chair)[11]
Analysis Committee: Jeffrey C Barrett[3], Paul R Burton[1], Dan Davison[11], Peter Donnelly[11], Doug Easton[12], David M Evans[3], Hin-Tak Leung[2], Jonathan L Marchini[11], Andrew P Morris[3], Chris CA Spencer[11], Martin D Tobin[1], Lon R Cardon[3,55], David G Clayton[2]
UK Blood Services & University of Cambridge Controls: Antony P Attwood[5,8], James P Boorman[8,9], Barbara Cant[8], Ursula Everson[13], Judith M Hussey[14], Jennifer D Jolley[8], Alexandra S Knight[8], Kerstin Koch[8], Elizabeth Meech[15], Sarah Nutland[2], Christopher V Prowse[16], Helen E Stevens[2], Niall C Taylor[8], Graham R Walters[17], Neil M Walker[2], Nicholas A Watkins[8,9], Thilo Winzer[8], John A Todd[2], Willem H Ouwehand[8,9]
1958 Birth Cohort Controls: Richard W Jones[18], Wendy L McArdle[18], Susan M Ring[18], David P Strachan[19], Marcus Pembrey[18,20]
Bipolar Disorder (Aberdeen): Gerome Breen[21], David St Clair[21]; (Birmingham): Sian Caesar[22], Katharine Gordon-Smith[22,23], Lisa Jones[22]; (Cardiff): Christine Fraser[23], Elaine K Green[23], Detelina Grozeva[23], Marian L Hamshere[23], Peter A Holmans[23], Ian R Jones[23], George Kirov[23], Valentina Moskivina[23], Ivan Nikolov[23], Michael C O'Donovan[23], Michael J Owen[23], Nick Craddock[23]; (London): David A Collier[24], Amanda Elkin[24], Anne Farmer[24], Richard Williamson[24], Peter McGuffin[24]; (Newcastle): Allan H Young[25], I Nicol Ferrier[25]
Coronary Artery Disease (Leeds): Stephen G Ball[26], Anthony J Balmforth[26], Jennifer H Barrett[26], Timothy D Bishop[26], Mark M Iles[26], Azhar Maqbool[26], Nadira Yuldasheva[26], Alistair S Hall[26]; (Leicester): Peter S Braund[10], Paul R Burton[1], Richard J Dixon[10], Massimo Mangino[10], Suzanne Stevens[10], Martin D Tobin[1], John R Thompson[1], Nilesh J Samani[10]
Crohn's Disease (Cambridge): Francesca Bredin[27], Mark Tremelling[27], Miles Parkes[27]; (Edinburgh): Hazel Drummond[28], Charles W Lees[28], Elaine R Nimmo[28], Jack Satsangi[28]; (London): Sheila A Fisher[29], Alastair Forbes[30], Cathryn M Lewis[29], Clive M Onnie[29], Natalie J Prescott[29], Jeremy Sanderson[31], Christopher G Matthew[29]; (Newcastle): Jamie Barbour[32], M Khalid Mohiuddin[32], Catherine E Todhunter[32], John C Mansfield[32]; (Oxford): Tariq Ahmad[33], Fraser R Cummings[33], Derek P Jewell[33]
Hypertension (Aberdeen): John Webster[34]; (Cambridge): Morris J Brown[35], David G Clayton[2]; (Evry, France): Mark G Lathrop[36]; (Glasgow): John Connell[37], Anna Dominiczak[37]; (Leicester): Nilesh J Samani[10]; (London): Carolina A Braga Marcano[38], Beverley Burke[38], Richard Dobson[38], Johannie Gungadoo[38], Kate L Lee[38], Patricia B Munroe[38], Stephen J Newhouse[38], Abiodun Onipinla[38], Chris Wallace[38], Mingzhan Xue[38], Mark Caulfield[38]; (Oxford): Martin Farrall[39]
Rheumatoid Arthritis: Anne Barton[40], The Biologics in RA Genetics and Genomics Study Syndicate (BRAGGS) Steering Committee*, Ian N Bruce[40], Hannah Donovan[40], Steve Eyre[40], Paul D Gilbert[40], Samantha L Hilder[40], Anne M Hinks[40], Sally L John[40], Catherine Potter[40], Alan J Silman[40], Deborah PM Symmons[40], Wendy Thomson[40], Jane Worthington[40]
Type 1 Diabetes: David G Clayton[2], David B Dunger[2,41], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], Barry Widmer[2,41], John A Todd[2]
Type 2 Diabetes (Exeter): Timothy M Frayling[42,43], Rachel M Freathy[42,43], Hana Lango[42,43], John R B Perry[42,43], Beverley M Shields[43], Michael N Weedon[42,43], Andrew T Hattersley[42,43]; (London): Graham A Hitman[44]; (Newcastle): Mark Walker[45]; (Oxford): Kate S Elliott[3,7], Christopher J Groves[7], Cecilia M Lindgren[3,7], Nigel W Rayner[3,7], Nicolas J Timpson[3,46], Eleftheria Zeggini[3,7], Mark I McCarthy[3,7]

Tuberculosis (Gambia): Melanie Newport[47], Giorgio Sirugo[47], (Oxford): Emily Lyons[3], Fredrik Vannberg[3], Adrian V S Hill[3]
Ankylosing Spondylitis: Linda A Bradbury[48], Claire Farrar[49], Jennifer J Pointon[49], Paul Wordsworth[49], Matthew A Brown[48,49]
Autoimmune Thyroid Disease: Jayne A Franklyn[50], Joanne M Heward[50], Matthew J Simmonds[50], Stephen CL Gough[50]
Breast Cancer: Sheila Seal[51], Breast Cancer Susceptibility Collaboration (UK)*, Michael R Stratton[51,52], Nazneen Rahman[51]
Multiple Sclerosis: Maria Ban[53], An Goris[53], Stephen J Sawcer[53], Alastair Compston[53]
Gambian Controls (Gambia): David Conway[47], Muminatou Jallow[47], Melanie Newport[47], Giorgio Sirugo[47]; (Oxford): Kirk A Rockett[3], Dominic P Kwiatkowski[3,5]
DNA, Genotyping, Data QC and Informatics (Wellcome Trust Sanger Institute, Hinxton): Suzannah J Bumpstead[5], Amy Chaney[5], Kate Downes[2,5], Mohammed JR Ghori[5], Rhian Gwilliam[5], Sarah E Hunt[5], Michael Inouye[5], Andrew Keniry[5], Emma King[5], Ralph McGinnis[5], Simon Potter[5], Rathi Ravindrarajah[5], Pamela Whittaker[5], Claire Widden[5], David Withers[5], Panos Deloukas[5]; (Cambridge): Hin-Tak Leung[2], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], John A Todd[2]
Statistics (Cambridge): Doug Easton[12], David G Clayton[2]; (Leicester): Paul R Burton[1], Martin D Tobin[1]; (Oxford): Jeffrey C Barrett[3], David M Evans[3], Andrew P Morris[3], Lon R Cardon[3,55]; (Oxford): Niall J Cardin[11], Dan Davison[11], Teresa Ferreira[11], Joanne Pereira-Gale[11], Ingeleif B Hallgrimsdóttir[11], Bryan N Howie[11], Jonathan L Marchini[11], Chris CA Spencer[11], Zhan Su[11], Yik Ying Teo[3,11], Damjan Vukcevic[11], Peter Donnelly[11]
Principal Investigators: David Bentley[5,54], Matthew A Brown[48,49], Lon R Cardon[3,55], Mark Caulfield[38], David G Clayton[2], Alastair Compston[53], Nick Craddock[23], Panos Deloukas[5], Peter Donnelly[11], Martin Farrall[39], Stephen CL Gough[50], Alistair S Hall[26], Andrew T Hattersley[42,43], Adrian V S Hill[3], Dominic P Kwiatkowski[3,5], Christopher G Matthew[29], Mark I McCarthy[3,7], Willem H Ouwehand[8,9], Miles Parkes[27], Marcus Pembrey[18,20], Nazneen Rahman[51], Nilesh J Samani[10], Michael R Stratton[51,52], John A Todd[2], Jane Worthington[40]
AITD Replication Group: Sarah L Mitchell[50], Paul R Newby[50], Oliver J Brand[50], Jackie Carr-Smith[50], Simon H S Pearce[56], Stephen C L Gough[50]
IL23R replication: R McGinnis[5], A Keniry[5], P Deloukas[5], TASC.
The Australo-Anglo-American Spondylitis Consortium (TASC): John D Reveille[57], Xiaodong Zhou[57], Linda A Bradbury[58], Anne-Marie Sims[58], Alison Dowling[58], Jacqueline Taylor[58], Tracy Doan[58], Lon R Cardon[55,59], John C Davis[60], Jennifer J Pointon[61], Laurie Savage[62], Michael M Ward[63], Thomas L Learch[64], Michael H Weisman[65], Paul Wordsworth[61], Matthew A Brown[58,61]

*See Supplementary Note for details.

Affiliations for participants: [1]Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK. [2]Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Cambridge CB2 0XY, UK. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. [4]Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK. [5]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [6]The Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK. [7]Oxford Centre for Diabetes, Endocrinology and Medicine, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK. [8]Department of Haematology, University of Cambridge, Long Road, Cambridge CB2 2PT, UK. [9]National Health Service Blood and Transplant, Cambridge Centre, Long Road, Cambridge CB2 2PT, UK. [10]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK. [11]Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK. [12]Cancer Research UK Genetic Epidemiology Unit, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK. [13]National Health Service Blood and Transplant, Sheffield Centre, Longley Lane, Sheffield S5 7JN, UK. [14]National Health Service Blood and Transplant, Brentwood Centre, Crescent Drive, Brentwood CM15 8DP, UK. [15]The Welsh Blood Service, Ely Valley Road, Talbot Green, Pontyclun CF72 9WB, UK. [16]The Scottish National Blood Transfusion Service, Ellen's Glen Road, Edinburgh EH17 7QT, UK. [17]National Health Service Blood and Transplant, Southampton Centre, Coxford Road, Southampton SO16 5AF, UK. [18]Avon Longitudinal Study of Parents and Children, University of Bristol, 24 Tyndall Avenue, Bristol BS8 1TQ, UK. [19]Division of Community Health Services, St. George's University of London, Crammer Terrace, London SW17 0RE, UK. [20]Institute of Child Health, University College London, 30 Guilford St., London WC1N 1EH, UK. [21]University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZD, UK. [22]Department of Psychiatry, Division of Neuroscience, Birmingham University, Birmingham B15 2QZ, UK. [23]Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff Univeristy, Heath Park, Cardiff CF14 4XN, UK. [24]SGDP, The Institute of Psychiatry, King's College London, De Crespigny Park, Denmark Hill, London SE5 8AF, UK. [25]School of Neurology, Neurobiology and Psychiatry, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne NE1 4LP, UK. [26]LIGHT and LIMM Research Institutes, Faculty of Medicine and Health, University of Leeds, Leeds LS1 3EX, UK. [27]IBD Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 2QQ, UK. [28]Gastrointestinal Unit, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU UK. [29]Department of Medical & Molecular Genetics, King's College London School of Medicine, 8th Floor Guy's Hospital, London SE1 9RT, UK. [30]Institute for Digestive Diseases, University College London Hospitals Trust, London NW1 2BU, UK. [31]Department of Gastroenterology, Guy's and St. Thomas' NHS Foundation Trust, London SE1 7EH, UK. [32]Department of Gastroenterology & Hepatology, University of Newcastle upon Tyne, Royal Victoria Infirmary, Newcastle upon Tyne NE1 4LP, UK. [33]Gastroenterology Unit, Radcliffe Infirmary, University of Oxford, Oxford OX2 6HE, UK. [34]Medicine and Therapeutics, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, Grampian AB9 2ZB, UK. [35]Clinical Pharmacology Unit and the Diabetes and Inflammation Laboratory, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK. [36]Centre National de Genotypage, 2 Rue Gaston Cremieux, Evry, Paris 91057, France. [37]BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow G12 8TA, UK. [38]Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London, Queen Mary's School of Medicine, Charterhouse Square, London EC1M 6BQ, UK. [39]Cardiovascular Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. [40]arc Epidemiology Research Unit, University of Manchester, Stopford Building, Oxford Rd, Manchester M13 9PT, UK. [41]Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 2QQ, UK. [42]Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, Magdalen Road, Exeter EX1 2LU, UK. [43]Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula Medical School, Barrack Road, Exeter EX2 5DU UK. [44]Centre for Diabetes and Metabolic Medicine, Barts and The London, Royal London Hospital, Whitechapel, London E1 1BB, UK. [45]Diabetes Research Group, School of Clinical Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK. [46]The MRC Centre for Causal Analyses in Translational Epidemiology, Bristol University, Canynge Hall, Whiteladies Rd., Bristol BS2 8PR, UK. [47]MRC Laboratories, Fajara, The Gambia. [48]Diamantina Institute for Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Woolloongabba, Queensland 4102, Australia. [49]Botnar Research Centre, University of Oxford, Headington, Oxford OX3 7BN, UK. [50]Department of Medicine, Division of Medical Sciences, Institute of Biomedical Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [51]Section of Cancer Genetics, Institute of Cancer Research, 15 Cotswold Road, Sutton SM2 5NG, UK. [52]Cancer Genome Project, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [53]Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK. [54]Illumina Cambridge, Chesterford Research Park, Little Chesterford, NR Saffron Walden, Essex CB10 1XL, UK. [55]Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. [56]University of Newcastle, Institute for Human Genetics, Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK. [57]Rheumatology and Clinical Immunogenetics, University of Texas-Houston Medical School, Houston, Texas 77030, USA. [58]Diamantina Institute for Cancer, Immunology and Metabolic Medicine, University of Queensland, Brisbane 4072, Australia. [59]Statistical Genetics, Wellcome Trust Centre for Human Genetics, Oxford OX37BN, UK. [60]Department of Rheumatology, University of California, San Francisco, California 94143, USA. [61]Botnar Research Centre, University of Oxford, Oxford OX37BN, UK. [62]The Spondylitis Association of America, Sherman Oaks, California 91403, USA. [63]National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [64]Department of Radiology and [65]Department of Medicine/Rheumatology, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics