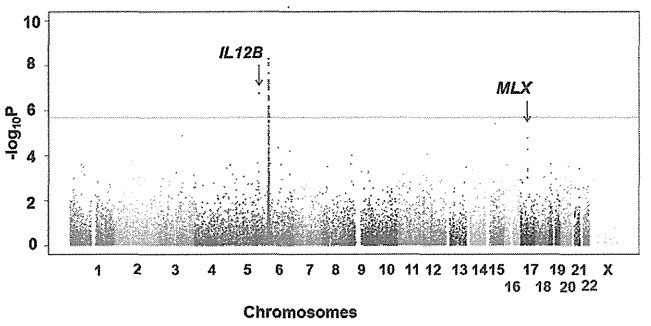


Figure 1. Manhattan Plot of Genome Scanning
The horizontal line indicates the significant level based on Bonferroni's correction. The HLA locus on chromosome 6 and the *IL12B* region on chromosome 5 reached the significant level.

in lymphoblastoid cells were obtained from GEO database (accession number GSE6536)[14] and analyzed for association with genotypes of rs6871626 obtained from HapMap project. Genevar software was used for analyzing the *IL12B* expression in adipose and skin in association with the rs6871626 genotypes.[15] Associations between genotypes and gene expression were evaluated by a linear regression analysis.

Associations between Genotypes and Clinical Phenotypes of TAK

Data of age at onset were analyzed for the association with the susceptibility alleles. AR, ischemic heart disease, and pulmonary infarction were selected for the association with genotypes as representative complications of TAK because cardiovascular event was the major cause of death in TAK individuals[16] and it was previously demonstrated that these phenotypes were associated with HLA-B*52:01,[17] suggesting that genetic backgrounds were at least partly responsible for these clinical manifestations. Data of the clinical manifestations were collected in Kyoto University Hospital or Tokyo Medical and Dental University by medical doctors who were blinded to genotype data reviewing clinical charts. Although AR evaluated by transthoracic echocardiography or angiography was positive for 44% of cases, other complications were found in less than 16%. Only AR was analyzed because of lack of power for other manifestations. Data for severity of AR assessed by the three categories[18] (mild, moderate, and severe) were also collected. C-reactive protein (CRP) was focused on as a biomarker reflecting disease activity. We calculated time-averaged CRP and dosage of prednisolone. Individuals who had visited hospitals for less than 500 days were excluded from the analysis of CRP. The associations between genotypes and clinical phenotypes were assessed by logistic regression analysis for existence of AR or linear regression analysis for severity of AR, time-averaged CRP, and age at onset. Time-averaged CRP was analyzed in condition with time-averaged dosage of prednisolone alone or in combination with rs3093059 genotypes in the *CRP* (MIM 123260) region. Associations between genotypes and clinical manifestations with p values less than 0.05 were regarded as significant.

Statistical Analysis

Statistical analyses were performed by PLINK v.1.07, R statistical software, or SPSS v.18.0.

Results

A summary of basic information of the subjects in our study is shown in Table 1. DNA samples from 167 cases and 663 healthy controls were genome scanned with the use of Illumina Human-Exome arrays containing 247,730 SNPs. One sample of the TAK cases and six samples in controls with success rates of less than 0.95 or with evidence of relatedness with other subjects (PI_HAT > 0.2 calculated by PLINK, see Subjects and Methods) were excluded from further analysis. The genotyping revealed that more than 80% of the markers in the array were monomorphic and 9% of the markers showed low minor allele frequency (<0.05) in the Japanese population, respectively. A total of 24,487 markers remained after filtering of SNPs that showed success rates of less than 0.95, deviation from HWE ($p < 1 \times 10^{-5}$) in either cases or controls, or minor allele frequencies of less than 0.05 in both cases and controls. The mean success rate of individuals was 0.999 after filtering.

Association studies were performed by chi-square test to compare allele frequencies between cases and controls. Population stratification was evaluated by QQ plot. The results indicated a lambda value of 1.05 in the QQ plot, indicating no excess population stratification in our study. Manhattan plot revealed that a region on chromosome 5 as well as the HLA locus showed significant associations that satisfied the genome-wide significant threshold obtained by Bonferroni's correction ($p = 2.0 \times 10^{-6}$; Figure 1). The associations were also confirmed by the imputed results (Figure S1 available online). rs4947248 in the *HLA-B* region, which is a known susceptibility gene to TAK, showed the strongest association ($p = 5.1 \times 10^{-9}$, OR = 2.17, 95% CI 1.67–2.82). rs9263739, a proxy of HLA-B*52:01 ($r^2 = 0.94$), similarly showed a significant association ($p = 8.0 \times 10^{-9}$, OR = 2.30, 95% CI 1.72–3.07; Table 2) and in moderate LD with rs4947248 ($D' = 0.95$, $r^2 = 0.58$). Because rs4947248 did not show evidence of an independent association from rs9263739 in logistic regression analysis ($p = 0.04$), we assumed that the top association in the HLA locus was attributable to HLA-B*52:01. rs6871626 in the *IL12B* region on chromosome 5 also showed a significant association ($p = 1.8 \times 10^{-7}$, OR = 1.90, 95% CI 1.49–2.42; Table 2 and Figure 2A). Four other loci showed suggestive associations in our study ($p < 5.0 \times 10^{-5}$; Table 2). No departure from HWE was observed for these six SNPs ($p \geq 0.041$).

A replication study was performed with the use of DNA samples from 212 cases and 1,322 controls. The six SNPs with p values less than 5.0×10^{-5} in the genome scanning were genotyped in the replication study. rs9263739 was selected as a representative of the associations in the HLA locus. As a result, the significant associations of TAK with rs6871626 and rs665268 in the *MLX* (MAX dimerization protein [MIM 602976]) region on chromosome 17 as well as rs9263739 were replicated ($p = 1.1 \times 10^{-7}$, 0.0032, and 6.0×10^{-15}, respectively; Table 2, Figures 2A

Table 2. Results of Association Studies for TAK Susceptibility

						Genome Scan			Replication			Meta-analysis	
SNP	Chr	Position	Gene	Ref(A1)	Var(A2)[a]	Case A2freq	Cont A2freq	p	Case A2freq	Cont A2freq	p	p	OR (95% CI)
rs10934853	3	129521063	*EEFSEC*	A	C	0.59	0.45	1.3×10^{-5}	0.52	0.47	0.066	2.6×10^{-5}	1.40 (1.20–1.64)
rs6871626	5	158759370	*IL12B*	C	A	0.53	0.37	1.8×10^{-7}	0.53	0.39	1.1×10^{-7}	1.7×10^{-13}	1.75 (1.42–2.16)
rs9263739	6	31219335	*CCHCR1*	C	T	0.27	0.14	8.0×10^{-9}	0.30	0.14	6.0×10^{-15}	2.8×10^{-21}	2.44 (2.03–2.93)
rs1570843	6	84577239	*RIPPLY2*	C	T	0.62	0.50	4.6×10^{-5}	0.54	0.51	0.19	3.1×10^{-4}	1.34 (1.14–1.57)
rs12102203	15	49578851	*DMXL2*	G	A	0.64	0.49	3.8×10^{-6}	0.53	0.54	0.71	0.0081	1.24 (1.06–1.46)
rs665268	17	37975555	*MLX*	A	G	0.58	0.44	1.7×10^{-5}	0.49	0.42	0.0032	5.2×10^{-7}	1.50 (1.28–1.76)

Abbreviations are as follows: chr, chromosome; ref, reference allele; var, variant allele; CaseA2freq, variant allele frequency in cases; ContA2freq, variant allele frequency in controls; OR, odds ratio; CI, confidence interval. Positions are according to National Center for Biotechnology Information (NCBI) build 36.
[a]Risk alleles for TAK based on the results of the genome scanning are set as variant alleles.

and 2B). The suggestive association on chromosome 15 (Figures S1 and 1) was not replicated. Again, no departure from HWE was observed ($p \geq 0.11$).

A combined study in which the associations in the two studies were integrated by inverse-variance method demonstrated that rs6871626, rs665268, and rs9263739 showed significant associations ($p = 1.7 \times 10^{-13}$, 5.2×10^{-7}, and 2.8×10^{-21}; OR = 1.75, 1.50, and 2.44; 95% CI 1.42–2.16, 1.28–1.76, and 2.03–2.93, respectively; Table 2) satisfying the significance obtained by Bonferroni's correction. rs6871626 and rs9263739 satisfied the more stringent, widely accepted genome-wide significance ($p = 5.0 \times 10^{-8}$).

Because it was suggested that genetic components had influence on the manifestations of the disease,[17] we analyzed whether the variant of the *IL12B* region had clinical effects on the disease course or severity. Age at onset was not associated with rs6871626 ($p = 0.36$), whereas a significant association between rs6871626 and development of AR was observed in a recessive model ($p = 0.0046$; Figure 3A). Focusing on the cases with AR, a significant association between rs6871626 and severity of AR was observed in the recessive model ($p = 0.0018$; Figure 3B). Risk allele of rs6871626 (A allele) also demonstrated a significant association with increased level of time-averaged CRP, which was a representative marker of the disease activity ($p = 0.021$; Figure 3C). The association between rs6871626 and CRP levels was independent from rs3093059 in the *CRP* region ($p = 0.029$), which showed the strongest association with circulating CRP levels in Japanese.[19] These associations between rs6871626 and clinical manifestations were independent from rs9263739 (conditioned p value of rs6871626 ≤ 0.020). Although rs665268 also demonstrated a significant association with development of AR in a dominant model ($p = 0.0089$; Figure S2A), the association was not significant in condition with rs9263739 ($p = 0.080$). No significant associations were observed between rs665268 and other clinical phenotypes (Figures S2B and S2C).

Next, we investigated the interaction between the *IL12B* and *HLA-B* loci to TAK susceptibility. The risk of TAK in the population positive for both rs6871626 A allele and rs9263739 T allele surpassed the product and sum of the risk in those who were positive for either rs6871626 A allele or rs9263739 T allele alone (Figure 4). The analysis revealed that those who were positive for both had OR of 6.00 (95% CI 4.22–8.55), whereas those who were positive for either rs9263739 T allele or rs6871626 A allele showed OR of 1.80 (95% CI 1.11–2.93) or 1.74 (95% CI 1.23–2.47), respectively. Interaction measures revealed RER of 3.46 ($p = 1.4 \times 10^{-5}$, 95% CI 1.90–5.01), AP of 0.58 ($p = 1.0 \times 10^{-12}$, 95% CI 0.42–0.73), and SI of 3.24 ($p = 0.00028$, 95% CI 1.72–6.11). This significant interaction between *IL12B* and *HLA-B* on TAK susceptibility could be observed in both studies (Table 3). The synergistic interaction effects between rs6871626 and rs9263739 were not evident in the clinical manifestations associated with rs6871626 (Figure S3). When we analyzed the interaction between the *MLX* and *HLA-B* regions, we observed suggestive interaction with RER of 1.73, AP of 0.43, and SI of 2.29 ($p \leq 0.027$; Figure S4 and Table S1). The associations between the interaction and clinical manifestations were not significant (Figure S5).

IL12B encodes a common subunit of the IL12 and IL23 protein, known as p40. Because previous studies showed that the *IL23R/IL12RB2* (MIM 607562/601642) region was associated with Behçet disease[20] (MIM 109650), another connective tissue disease where vasculitis is involved in its pathology, we investigated this region for the possible associations in the current study. As a result, no suggestive association was found, either in our study or in the imputed results (Figure S6).

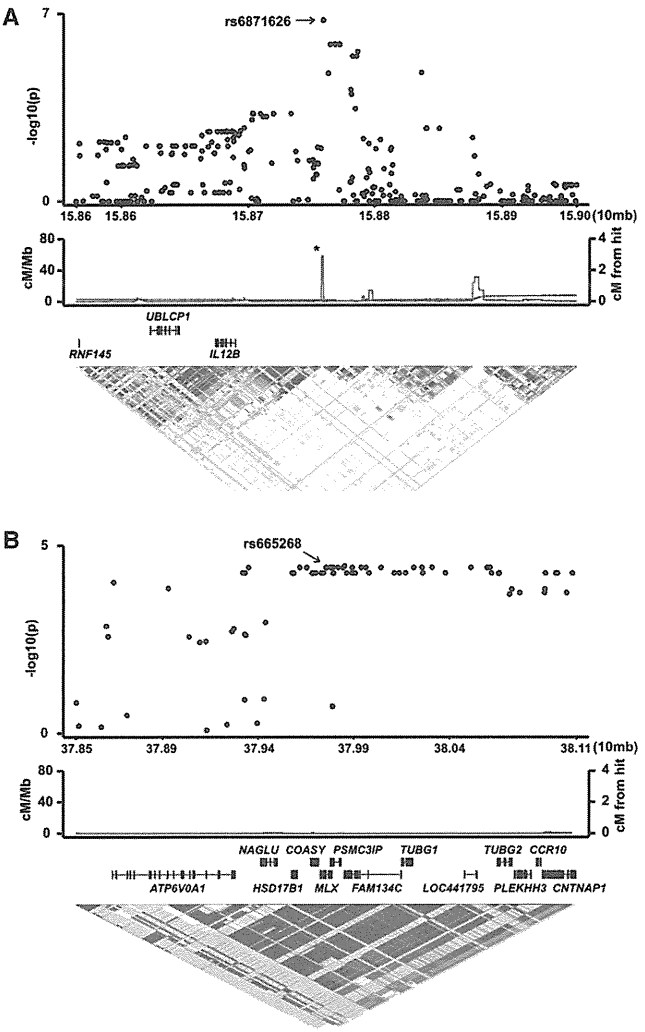


Figure 2. Associations of the *IL12B* and *MLX* Regions with the Susceptibility to TAK

Associations of SNPs in the (A) *IL12B* and (B) *MLX* regions in the genome scanning are plotted according to the position of the markers. Red circles indicate results of the current genome scanning. Blue circles indicate results of the imputation analysis based on the current results. The middle panel indicates recombination rates. The lower panel indicates LD of markers. Asterisk indicates a recombination hot spot in the *IL12B* region.

Discussion

This study provides a convincing evidence of associations between non-HLA genes and TAK susceptibility along with a synergistic role of susceptibility genes to TAK. The lack of evidence for associations of non-HLA genes with TAK so far is attributable to the lack of GWASs of TAK performed to date. Low prevalence of this disease had made it difficult to collect DNA samples to obtain sufficient power to detect susceptibility genes and perform a GWAS. Previous studies have revealed that the *IL12B* region was associated with a wide variety of autoimmune disorders and infectious diseases, including psoriasis[21–23] (MIM 177900), ankylosing spondylitis[24] (MIM 106300), Crohn disease[25] (CD [MIM 266600]), ulcerative colitis[26] (UC [MIM 191390]), and leprosy[27] (MIM 609888). rs6871626 showed a significant association with UC and leprosy over the genome-wide significance. Notably, rs6871626 A allele is susceptible to UC but protective against leprosy. A previous study from Turkey reported a suggestive association of TAK with rs3212227 in the 3′ UTR of the *IL12B* region.[28] rs3212227 is not in strong LD with rs6871626 in the Japanese population ($r^2 = 0.11$) and in Europeans ($r^2 = 0.06$) because of a recombination hot spot adjacent to rs6871626 (Figure 2A). In fact, an imputed association of rs3212227 with TAK in the current study resulted in only a suggestive association ($p = 0.0027$). There is a possibility that rs6871626 was responsible for the suggestive association between rs3212227 and TAK reported in the Turkish population. The association between gene expression and SNPs in the *IL12B* region appears to be complicated and inconsistent across different studies. rs3212227 in the 3′ UTR and rs17860508, an ins/del polymorphism in the promoter region of *IL12B*, were shown to have potential effects on the gene expression.[29,30] However, the previous studies showed that the association patterns varied according to the cell type and the protocol used for stimulation.[31–33] No previous report analyzed the effects of rs6871626 on the gene expression of *IL12B*. Although our in silico analysis failed to show the effects of rs6871626 on *IL12B* expression (data not shown, see Subjects and Methods), specific cell types or stimulus could lead to a significant association. Because a recent study showed that a haplotype of SNPs in the *IL12B* region could influence the gene and protein expression of *IL12B*,[22] a combination of rs6871626 and other SNPs in the *IL12B* region might lead to consistent results.

The associations between rs6871626 and clinical manifestations of TAK suggest the fundamental effects of IL-12p40 protein on TAK progression as well as TAK onset. We found that HLA-B*52:01 was associated with AR as reported previously ($p = 0.00014$). This finding supported the accuracy of our data. Although the risk allele of rs6871626 was associated with a significant dose-dependent increase in risk and severity of AR and in circulating CRP levels ($p = 0.013$, 0.030, and 0.023, respectively), these associations were more evident in a recessive manner. This raises a possibility that those who are homozygous for rs6871626 have strong disease activity that exceeds the additive disease activity of cases with single risk alleles, leading to severe destruction of aortic valve. Genetic variations in *IL12B* are known to influence the risk of psoriasis[21–23] and CD.[25] Because ustekinumab, a monoclonal antibody against IL-12p40, is an effective treatment for both diseases,[34,35] our findings would raise a possibility of its therapeutic use for TAK by targeting the IL-12/23 pathway. A previous study reported that the level of IL-12 protein was increased in TAK cases compared to healthy populations,[36] whereas there have been no reports addressing the circulating levels of IL-23 in TAK cases. IL-12 directly leads to type 1 helper T cell proliferation[37] and IL-23 upregulates IL-17 production and

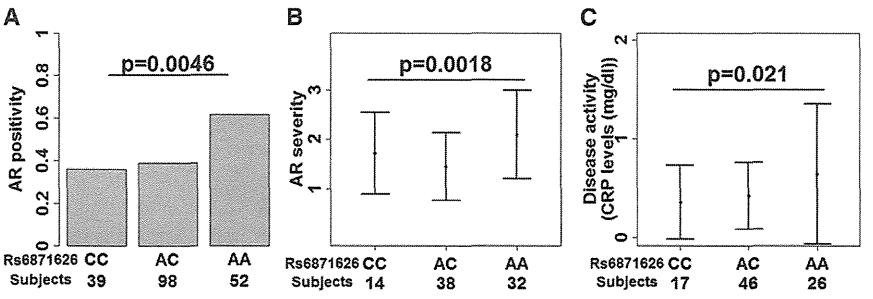


Figure 3. Associations between rs6871626 Genotypes and Clinical Manifestations of TAK

An association between rs6871626 genotypes and (A) development of AR, (B) severity of AR, and (C) time-averaged CRP levels in TAK cases. The p value was calculated by (A) logistic regression analysis, (B) linear regression analysis, and (C) linear regression analysis with time-averaged dosage of prednisolone as covariate. The recessive model is applied to all calculations. Severity of 1 to 3 in AR corresponds to mild, moderate, and severe, respectively. Mean ± SD are indicated for (B) and (C).

supports survival of activated Th17 cells.[38] Further analyses addressing circulating T cells in individuals with TAK or cell types infiltrating the artery specimen obtained from cases would provide clues to specify a critical pathway in TAK pathology.

The synergistic effect between rs6871626 and HLA-B*52:01 was notable. Those carrying both risk alleles had OR of 6.00 in comparison with those not carrying any risk alleles. Combination of rs6871626 and HLA-B*52:01 showed tendency of severe clinical phenotypes. Thus, we assume that increase of subjects and extraction of subjects who are homozygous for rs6871626 and positive for HLA-B*52:01 would provide evidence for significant effects of the combination on the disease phenotypes. The synergistic effect of these two loci raises a possibility that immune-related cells that recognize a yet-to-be-determined antigen through HLA-B*52:01 can be overactivated by IL-12/23 whose expression or function is modulated by rs6871626. In vitro analysis of immune-related cells from cases with TAK or healthy individuals would provide functional evidence of this synergistic role in the TAK pathogenesis.

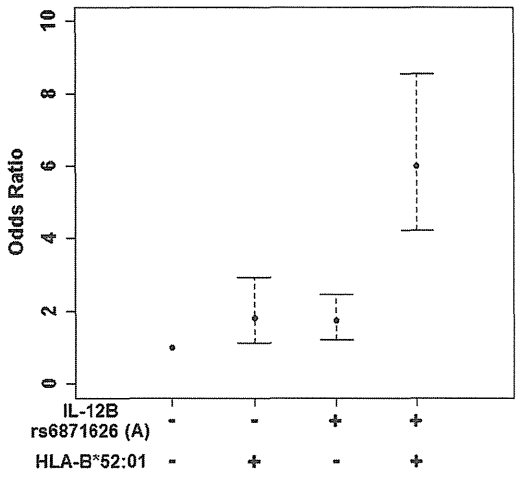


Figure 4. A Synergistic Effect between *IL12B* and HLA-B*52:01 on TAK Susceptibility

ORs are shown for the four strata of subjects according to combination of rs6871626 and rs9263739 genotypes. Those who are negative for rs9263739 T allele, a proxy of HLA-B*52:01, and rs6871626 A allele are used as reference. ORs and 95% CI are indicated.

rs665268 is a missense mutation of *MLX* that alters the 139th glutamine to arginine (Gln139Arg). MLX is a member of the basic helix-loop-helix leucine zipper (bHLH-Zip) transcription factor family and regulates gene expression by forming heterodimers with Mad protein.[39] The 17q21 region, whose associations with other autoimmune diseases including psoriasis[40] and CD[41] were shown, contains a number of genes including immune-related genes and polymorphisms that are in strong LD with each other (Figure 2B), so the corresponding gene to TAK susceptibility was inconclusive. Because risk allele frequency of rs665268 is comparable to that of rs6871626, the lack of associations between rs665268 and clinical manifestations and the weaker interaction between rs665268 and HLA-B*52:01 compared to rs6871626 might be a reflection of the milder effect of rs665268 on TAK progression. No interaction was observed between rs665268 and rs6871626 (data not shown).

We set the relatively low cut-off value of imputation score (Rsq ≥ 0.3) in the imputation analysis to increase sensitivity at the expense of specificity, but we failed to find other candidates of susceptibility loci. Another imputation analysis based on the data from the 1000 Genomes Project[42] revealed the same signals as the current study (data not shown). However, because the array used in the current study focused on SNPs in exons or nearby genes, it did not fully cover the whole genome with dense markers even in imputation analysis. There is a possibility that other SNPs not tagged by the markers on the array are associated with TAK. When the associations in the HLA locus were conditioned by rs9263739 or rs4947248, the most significantly associated SNPs, suggestive association signals in this locus could still be observed (the smallest p value = 5.5×10^{-5}, data not shown). The use of arrays with denser markers especially in intergene regions and using an increased number of cases could lead to the discovery of other susceptibility regions or independent associations in the HLA locus. Considering that both of the non-HLA susceptibility loci to TAK found in the current study are also associated with psoriasis and inflammatory bowel diseases, further analysis of TAK susceptibility genes would reveal other overlapping loci and common autoimmune mechanisms between TAK and other autoimmune diseases. It is feasible to

Table 3. Synergistic Effects between *IL12B* and HLA-B*52:01 in Each Study

	RERI		AP		SI	
Study	**(95% CI)**	**p**	**(95% CI)**	**p**	**(95% CI)**	**p**
Genome scanning	2.90 (0.60–5.20)	0.014	0.50 (0.23–0.78)	0.00034	2.57 (1.08–6.09)	0.032
Replication study	3.87 (1.70–6.05)	0.00049	0.62 (0.42–0.81)	4.7×10^{-10}	3.76 (1.51–9.32)	0.0043
Combined study	3.46 (1.90–5.02)	1.4×10^{-5}	0.58 (0.42–0.73)	1.0×10^{-12}	3.24 (1.72–6.11)	0.00028

Abbreviations are as follows: RERI, relative excess risk; AP, attributable proportion; SI, synergy index; CI, confidence interval.

analyze whether these two loci are associated with TAK and whether the interactions are observed in other populations.

Taken together, the current study identified two susceptibility genes to TAK and provided evidence of a common immunological pathway exerted by the *IL12B* region that is involved in the etiology of TAK and other autoimmune disorders and of its synergistic role with HLA in the susceptibility to TAK.

Supplemental Data

Supplemental Data include six figures and one table and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

We'd like to thank all the individuals with TAK who gave their blood samples and medical staffs to help us for this study; Miki Kokubo for her excellent technique to extract DNA; Kayo Umemoto for coordination of meetings to obtain blood samples; and Masashi Akizuki for his help to collect DNA samples. This study is supported by Kyoto University Step-up grant, Grants-in-aid from Research on rare and intractable diseases, the Ministry of Health, Labor, and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, grants from SENSHIN Medical Research Foundation (to T. Matsumura), and the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program) from the Japan Society for the Promotion of Science (to R.N.).

Received: March 23, 2013
Revised: May 18, 2013
Accepted: May 28, 2013
Published: July 3, 2013

Web Resources

The URLs for data presented herein are as follows:

Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo/
Genevar (Gene Expression Variation), http://www.sanger.ac.uk/resources/software/genevar/
International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/
R statistical software, http://www.r-project.org/

References

1. Isobe, M. (2013). Takayasu arteritis revisited: Current diagnosis and treatment. Int. J. Cardiol. Published online February 13, 2013. http://dx.doi.org/10.1016/j.ijcard.2013.01.022.
2. Dreyer, L., Faurschou, M., and Baslund, B. (2011). A population-based study of Takayasu's arteritis in eastern Denmark. Clin. Exp. Rheumatol. *29*(1, Suppl 64), S40–S42.
3. Jennette, J.C., Falk, R.J., Bacon, P.A., Basu, N., Cid, M.C., Ferrario, F., Flores-Suarez, L.F., Gross, W.L., Guillevin, L., Hagen, E.C., et al. (2013). 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. Arthritis Rheum. *65*, 1–11.
4. Yoshida, M., Kimura, A., Katsuragi, K., Numano, F., and Sasazuki, T. (1993). DNA typing of HLA-B gene in Takayasu's arteritis. Tissue Antigens *42*, 87–90.
5. Yajima, M., Numano, F., Park, Y.B., and Sagar, S. (1994). Comparative studies of patients with Takayasu arteritis in Japan, Korea and India—comparison of clinical manifestations, angiography and HLA-B antigen. Jpn. Circ. J. *58*, 9–14.
6. Sahin, Z., Bıcakcıgil, M., Aksu, K., Kamali, S., Akar, S., Onen, F., Karadag, O., Ozbalkan, Z., Ates, A., Ozer, H.T., et al.; Turkish Takayasu Study Group. (2012). Takayasu's arteritis is associated with HLA-B*52, but not with HLA-B*51, in Turkey. Arthritis Res. Ther. *14*, R27.
7. Vargas-Alarcón, G., Flores-Domínguez, C., Hernández-Pacheco, G., Zuñiga, J., Gamboa, R., Soto, M.E., Granados, J., and Reyes, P.A. (2001). Immunogenetics and clinical aspects of Takayasu's arteritis patients in a Mexican Mestizo population. Clin. Exp. Rheumatol. *19*, 439–443.
8. Arend, W.P., Michel, B.A., Bloch, D.A., Hunder, G.G., Calabrese, L.H., Edworthy, S.M., Fauci, A.S., Leavitt, R.Y., Lie, J.T., Lightfoot, R.W., Jr., et al. (1990). The American College of Rheumatology 1990 criteria for the classification of Takayasu arteritis. Arthritis Rheum. *33*, 1129–1134.
9. JCS Joint Working Group; Japanese Circulation Society. (2011). Guideline for management of vasculitis syndrome (JCS 2008). Circ. J. *75*, 474–503.
10. Yoshimura, K., Nakayama, T., Sekine, A., Matsuda, F., Kosugi, S., Sugino, Y., Yoshimura, K., and Ogawa, O.; Nagahama Cohort Research Group. (2013). Prevalence of postmicturition urinary incontinence in Japanese men: Comparison with other types of incontinence. Int. J. Urol. Published online January 10, 2013. http://dx.doi.org/10.1111/iju.12074.
11. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

12. Li, Y., Willer, C.J., Ding, J., Scheet, P., and Abecasis, G.R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. *34*, 816–834.
13. Andersson, T., Alfredsson, L., Källberg, H., Zdravkovic, S., and Ahlbom, A. (2005). Calculating measures of biological interaction. Eur. J. Epidemiol. *20*, 575–579.
14. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science *315*, 848–853.
15. Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P., and Dermitzakis, E.T. (2010). Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics *26*, 2474–2476.
16. Andrews, J., and Mason, J.C. (2007). Takayasu's arteritis—recent advances in imaging offer promise. Rheumatology (Oxford) *46*, 6–15.
17. Kitamura, H., Kobayashi, Y., Kimura, A., and Numano, F. (1998). Association of clinical manifestations with HLA-B alleles in Takayasu arteritis. Int. J. Cardiol. *66(Suppl 1)*, S121–S126.
18. Warnes, C.A., Williams, R.G., Bashore, T.M., Child, J.S., Connolly, H.M., Dearani, J.A., del Nido, P., Fasules, J.W., Graham, T.P., Jr., Hijazi, Z.M., et al.; American College of Cardiology; American Heart Association Task Force on Practice Guidelines (Writing Committee to Develop Guidelines on the Management of Adults With Congenital Heart Disease); American Society of Echocardiography; Heart Rhythm Society; International Society for Adult Congenital Heart Disease; Society for Cardiovascular Angiography and Interventions; Society of Thoracic Surgeons. (2008). ACC/AHA 2008 guidelines for the management of adults with congenital heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Develop Guidelines on the Management of Adults With Congenital Heart Disease). Developed in Collaboration With the American Society of Echocardiography, Heart Rhythm Society, International Society for Adult Congenital Heart Disease, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J. Am. Coll. Cardiol. *52*, e143–e263.
19. Okada, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Kamatani, Y., Hosono, N., Tsunoda, T., Matsuda, K., Tanaka, T., Kubo, M., et al. (2011). Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. Hum. Mol. Genet. *20*, 1224–1231.
20. Mizuki, N., Meguro, A., Ota, M., Ohno, S., Shiota, T., Kawagoe, T., Ito, N., Kera, J., Okada, E., Yatsu, K., et al. (2010). Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behçet's disease susceptibility loci. Nat. Genet. *42*, 703–706.
21. Tsunemi, Y., Saeki, H., Nakamura, K., Sekiya, T., Hirai, K., Fujita, H., Asano, N., Kishimoto, M., Tanida, Y., Kakinuma, T., et al. (2002). Interleukin-12 p40 gene (IL12B) 3′-untranslated region polymorphism is associated with susceptibility to atopic dermatitis and psoriasis vulgaris. J. Dermatol. Sci. *30*, 161–166.
22. Johnston, A., Xing, X., Swindell, W.R., Kochkodan, J., Riblett, M., Nair, R.P., Stuart, P.E., Ding, J., Voorhees, J.J., Elder, J.T., and Gudjonsson, J.E. (2013). Susceptibility-associated genetic variation at IL12B enhances Th1 polarization in psoriasis. Hum. Mol. Genet. *22*, 1807–1815.
23. Cargill, M., Schrodi, S.J., Chang, M., Garcia, V.E., Brandon, R., Callis, K.P., Matsunami, N., Ardlie, K.G., Civello, D., Catanese, J.J., et al. (2007). 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.
24. Evans, D.M., Spencer, C.C., Pointon, J.J., Su, Z., Harvey, D., Kochan, G., Oppermann, U., Dilthey, A., Pirinen, M., Stone, M.A., et al.; Spondyloarthritis Research Consortium of Canada (SPARCC); Australo-Anglo-American Spondyloarthritis Consortium (TASC); Wellcome Trust Case Control Consortium 2 (WTCCC2). (2011). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat. Genet. *43*, 761–767.
25. Parkes, M., Barrett, J.C., Prescott, N.J., Tremelling, M., Anderson, C.A., Fisher, S.A., Roberts, R.G., Nimmo, E.R., Cummings, F.R., Soars, D., et al.; Wellcome Trust Case Control Consortium. (2007). Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat. Genet. *39*, 830–832.
26. Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D'Amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A., et al. (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat. Genet. *43*, 246–252.
27. Liu, H., Irwanto, A., Tian, H., Fu, X., Yu, Y., Yu, G., Low, H., Chu, T., Li, Y., Shi, B., et al. (2012). Identification of IL18-RAP/IL18R1 and IL12B as leprosy risk genes demonstrates shared pathogenesis between inflammation and infectious diseases. Am. J. Hum. Genet. *91*, 935–941.
28. Saruhan-Direskeneli, G., Biçakçigil, M., Yilmaz, V., Kamali, S., Aksu, K., Fresko, I., Akkoç, N., Kiraz, S., Ozer, H.T., Tunç, E., et al.; Rheumatology Education and Research Society Vasculitis Study Group. (2006). Interleukin (IL)-12, IL-2, and IL-6 gene polymorphisms in Takayasu's arteritis from Turkey. Hum. Immunol. *67*, 735–740.
29. Morahan, G., Huang, D., Wu, M., Holt, B.J., White, G.P., Kendall, G.E., Sly, P.D., and Holt, P.G. (2002). Association of IL12B promoter polymorphism with severity of atopic and non-atopic asthma in children. Lancet *360*, 455–459.
30. Stanilova, S., and Miteva, L. (2005). Taq-I polymorphism in 3′UTR of the IL-12B and association with IL-12p40 production from human PBMC. Genes Immun. *6*, 364–366.
31. Basu, M., Das, T., Ghosh, A., Majumder, S., Maji, A.K., Kanjilal, S.D., Mukhopadhyay, I., Roychowdhury, S., Banerjee, S., and Sengupta, S. (2012). Gene-gene interaction and functional impact of polymorphisms on innate immune genes in controlling *Plasmodium falciparum* blood infection level. PLoS ONE *7*, e46441.
32. Yanagihori, H., Oyama, N., Nakamura, K., Mizuki, N., Oguma, K., and Kaneko, F. (2006). Role of IL-12B promoter polymorphism in Adamantiades-Behcet's disease susceptibility: An involvement of Th1 immunoreactivity against *Streptococcus Sanguinis* antigen. J. Invest. Dermatol. *126*, 1534–1540.
33. Yilmaz, V., Yentür, S.P., and Saruhan-Direskeneli, G. (2005). IL-12 and IL-10 polymorphisms and their effects on cytokine production. Cytokine *30*, 188–194.
34. Yeilding, N., Szapary, P., Brodmerkel, C., Benson, J., Plotnick, M., Zhou, H., Goyal, K., Schenkel, B., Giles-Komar, J., Mascelli, M.A., and Guzzo, C. (2012). Development of the IL-12/23

antagonist ustekinumab in psoriasis: past, present, and future perspectives—an update. Ann. N Y Acad. Sci. *1263*, 1–12.
35. Sandborn, W.J., Gasink, C., Gao, L.L., Blank, M.A., Johanns, J., Guzzo, C., Sands, B.E., Hanauer, S.B., Targan, S., Rutgeerts, P., et al.; CERTIFI Study Group. (2012). Ustekinumab induction and maintenance therapy in refractory Crohn's disease. N. Engl. J. Med. *367*, 1519–1528.
36. Verma, D.K., Tripathy, N.K., Verma, N.S., and Tiwari, S. (2005). Interleukin 12 in Takayasu's arteritis: plasma concentrations and relationship with disease activity. J. Rheumatol. *32*, 2361–2363.
37. Murphy, K.M., and Reiner, S.L. (2002). The lineage decisions of helper T cells. Nat. Rev. Immunol. *2*, 933–944.
38. Bettelli, E., Carrier, Y., Gao, W., Korn, T., Strom, T.B., Oukka, M., Weiner, H.L., and Kuchroo, V.K. (2006). Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. Nature *441*, 235–238.
39. Cairo, S., Merla, G., Urbinati, F., Ballabio, A., and Reymond, A. (2001). WBSCR14, a gene mapping to the Williams-Beuren syndrome deleted region, is a new member of the Mlx transcription factor network. Hum. Mol. Genet. *10*, 617–627.
40. Tsoi, L.C., Spain, S.L., Knight, J., Ellinghaus, E., Stuart, P.E., Capon, F., Ding, J., Li, Y., Tejasvi, T., Gudjonsson, J.E., et al.; Collaborative Association Study of Psoriasis (CASP); Genetic Analysis of Psoriasis Consortium; Psoriasis Association Genetics Extension; Wellcome Trust Case Control Consortium 2. (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat. Genet. *44*, 1341–1348.
41. Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat. Genet. *42*, 1118–1125.
42. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A.; 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

OPEN

Crowdsourcing genetic prediction of clinical utility in the Rheumatoid Arthritis Responder Challenge

To the Editor:

The ability to translate large-scale genetics and genomics data into biological knowledge has not kept pace with our ability to generate these data sets. As a consequence, a major bottleneck in biomedical research has become access to data within a computational workspace that allows for robust, collaborative analyses. One innovative solution is to bring together scientific data, code, tools and disease models into an open commons or workspace, for example, the Synapse platform of Sage Bionetworks[1]. This environment allows for real-time sharing of large genomic data sets, continuous peer review and rapid learning within a system constructed to provide data access in a manner aligned with the informed consent provided by patients and research participants.

This crowdsourcing approach has been used to predict breast cancer survival from clinical and omics data[2] and was suggested as a way to find new drugs[3] by soliciting contributions from a large online community collaborating or competing to answer an inherently difficult but important question[4]. Researchers initiating an open challenge invite solutions but also incentivize the process by offering new data, a process in which the participants' methods can be assessed by testing their predictions against previously unseen data sets. This year, Sage and DREAM (Dialogue for Reverse Engineering Assessments and Methods) are running four open challenges (http://www.sagebase.org/challenges-overview/2013-dream-challenges/).

Here we announce the challenge to develop genetic predictors of response to immunosuppressive therapy in a common autoimmune disease, rheumatoid arthritis (RA). Disease-modifying antirheumatic drugs such as those that block the inflammatory cytokine tumor necrosis factor-α (known as anti-TNF therapy) are not effective in all patients with RA, with up to one-third of such patients failing to enter clinical remission after a standard course of therapy[5]. Moreover, the biological mechanisms underlying this failure are unknown, limiting the development of clinical biomarkers to guide either this therapy or the development of new drugs to target refractory cases.

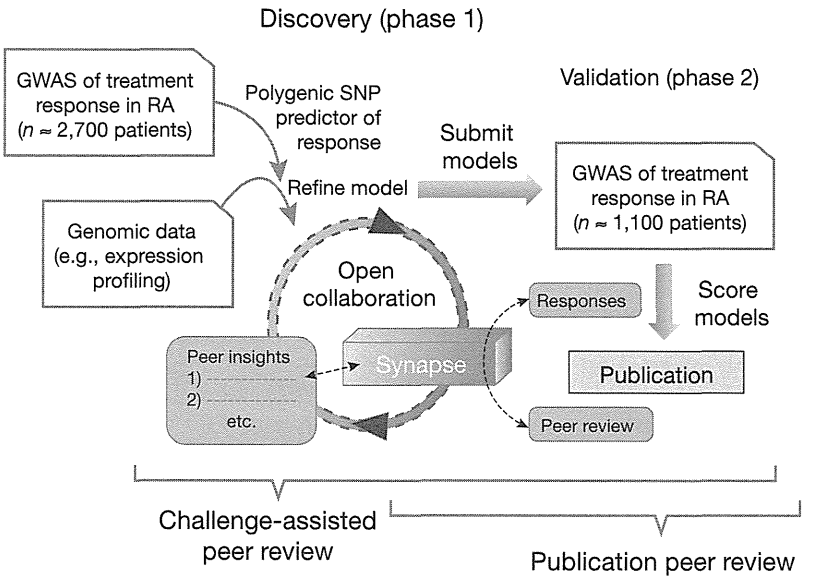


Figure 1 Overview of the Rheumatoid Arthritis Responder Challenge. There are two phases to the challenge. In phase 1 (discovery), analysts build genetic models of response to anti-TNF therapy using SNP data from a GWAS of ~2,700 patients with RA. To facilitate model building, additional genomic data will be made available. In a model of open collaboration, participants will use Synapse to post code, share insights and engage in rapid learning prepublication. In phase 2 (validation), models will be posted, tested and scored in an independent GWAS data set of ~1,100 patients with RA treated with anti-TNF therapy. To complement challenge-assisted peer review (which occurs in both the discovery and validation phases), conventional peer review will have access to Synapse to understand the iterative process of model building. Synapse will allow study investigators to respond to peer-review critiques and resubmit versions of their models and studies.

The Rheumatoid Arthritis Responder Challenge is for teams to build the best genetic predictor of response to anti-TNF therapy. There are two phases to the challenge: discovery and validation (**Fig. 1**). In the discovery phase, teams will utilize genomic data sets—several of which will be generated for the purposes of this challenge—and a variety of analytical methods to build predictive polygenic models of treatment response. We recently published a genome-wide association study (GWAS) in ~2,700 patients with RA treated with anti-TNF therapy[6]. Our GWAS data indicate that the genetic architecture of the anti-TNF response is probably highly polygenic, similar to what has been observed for other complex traits, such as risk of RA[7]. Importantly, our challenge will incorporate a new GWAS data set, which will be used in the validation phase, in which models built in the discovery phase are tested. The data set of ~1,100 patients with RA treated with anti-TNF therapy will be made available though

a public-private partnership between the Consortium of Rheumatology Researchers of North America, Inc. (CORRONA) and the Pharmacogenomics Research Network (PGRN) sponsored by the National Institute of General Medical Sciences (NIGMS) and the US National Institutes of Health (NIH).

A unique component of our Rheumatoid Arthritis Responder Challenge is the diversity of participation across a number of groups from academic institutions, private foundations and for-profit companies. In addition to support from CORRONA and PGRN, we received funding from pharmaceutical companies (see complete list on our website; link below) and a private foundation (the Arthritis Foundation) to support the public commons. We also received support from the Arthritis Internet Registry (AIR) and the Broad Institute to generate new genomic data sets, as well as in-kind support from a large number of academic collaborators from across the world to make GWAS data available in the discovery phase. We anticipate that a winning classifier could enable a follow-on prospective clinical trial within the group of appropriately consented patients in AIR.

Through Synapse, analysts who are inclined to establish collaborations will have the opportunity to see in real time the models that others are using so that each team can learn from the others (**Fig. 1**). A leaderboard will show the relative performance ranking of the different teams on the basis of a crossvalidation strategy designed to minimize overfitting. During the discovery phase, teams that choose to collaborate with each other will have the opportunity to check each other's algorithms for readability, speed and reproducibility. Then, during the validation phase, each team will submit computer code, which the Sage-DREAM team (http://www.sagebase.org/) will test in Synapse to establish whether it runs as expected to predict if a subject is an anti-TNF therapy responder or nonresponder on the basis of the GWAS data. Predefined performance metrics will be used to objectively determine the accuracy of the predictions, their statistical significance and the final performance ranking of the participating teams. The team that develops the most highly predictive model will be deemed the 'winner', with precise attribution of contributor roles going to all members of teams that contributed to building the final consensus model.

The best-performing models, therefore, will have passed a test of performance that is outside the realm of, and complements, traditional peer review. Indeed, this stringent test of method performance can be used as an enhanced way of publication vetting, what we call 'challenge-assisted peer review'. Traditional peer review is essential for ensuring the clarity, originality, contextualization and logical thread of a discrete set of work that is ready to be used by researchers in the form of a published article. However, the complexity of working with omics data—entailing multiple analytical decisions, computational simulations and statistical calculations—means that referees are challenged to follow and check the components of even a traditional research paper. In our Rheumatoid Arthritis Responder Challenge, we will explore the feasibility of enhancing the reliability and transparency of conventional peer review in partnership with *Nature Genetics*. This can be achieved if the referees and authors of the paper reporting on the best-performing methods in the challenge are willing to leave their comments openly (yet anonymously) on the Synapse platform (**Fig. 1**). We anticipate that the challenge-based assessment of accuracy will provide an objective metric of performance and a comparison with state-of-the-art analytical methodologies that will greatly enhance the task of refereeing a body of work with more quality control than is currently provided by conventional peer review.

In conclusion, we believe that the Rheumatoid Arthritis Responder Challenge is an apt use of crowdsourcing in human genetics to gain insight into clinical prediction and disease biology. Details of the challenge, including the rules by which the models will be judged, can be found at https://synapse.prod.sagebase.org/#!Synapse:syn1734172.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Robert M Plenge[1,2], Jeffrey D Greenberg[3], Lara M Mangravite[4], Jonathan M J Derry[4], Eli A Stahl[5], Marieke J H Coenen[6], Anne Barton[7], Leonid Padyukov[8], Lars Klareskog[8], Peter K Gregersen[9], Xavier Mariette[10], Larry W Moreland[11], S Louis Bridges Jr[12], Niek de Vries[13], Tom W J Huizinga[14], Henk-Jan Guchelaar[15], International Rheumatoid Arthritis Consortium (INTERACT)[16], Stephen H Friend[4,17] & Gustavo Stolovitzky[17,18].

[1]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [2]Medical and Population Genetics Program, Chemical Biology Program, Broad Institute, Cambridge, Massachusetts, USA. [3]Division of Rheumatology, New York University School of Medicine, New York, New York, USA. [4]Sage Bionetworks, Seattle, Washington, USA. [5]Division of Psychiatric Genomics and Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mt. Sinai, New York, New York, USA. [6]Department of Human Genetics, Radboud University Medical Centre, Nijmegen, The Netherlands. [7]Arthritis Research UK Epidemiology Unit, Musculoskeletal Research Group, University of Manchester, Manchester Academic Health Sciences Centre, Manchester, UK. [8]Rheumatology Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden. [9]The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, USA. [10]Université Paris-Sud, Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpitaux Universitaires Paris-Sud, Institut National de la Santé et de la Recherche Médicale (INSERM) U1012, Rheumatology, le Kremlin-Bicêtre, France. [11]Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [12]Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, Alabama, USA. [13]Department of Clinical Immunology and Rheumatology, Academic Medical Center (AMC), University of Amsterdam, Amsterdam, The Netherlands. [14]Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands. [15]Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, The Netherlands. [16]A full list of members is provided at https://synapse.prod.sagebase.org/#!Synapse:syn1734172. [17]Sage/DREAM Project, Sage Bionetworks, Seattle, Washington, USA. [18]IBM Computational Biology Center, T.J. Watson Research Center, Yorktown Heights, New York, USA.
e-mail: rplenge@partners.org.

1. Derry, J.M.J. *et al. Nat. Genet.* **44**, 127–130 (2012).
2. Margolin, A.A. *et al. Sci. Transl. Med.* **5**, 181re1 (2013).
3. Wadman, M. *Nature* **490**, 15 (2012).
4. Costello, J.C. & Stolovitzky, G. *Clin. Pharmacol. Ther.* **93**, 396–398 (2013)
5. McInnes, I.B. & Schett, G. *N. Engl. J. Med.* **365**, 2205–2219 (2011).
6. Cui, J. *et al. PLoS Genet.* published online, http://dx.doi.org/10.1371/journal.pgen.1003394 (28 March 2013).
7. Stahl, E.A. *et al. Nat. Genet.* **44**, 483–489 (2012).

PLOS | GENETICS

Genome-Wide Association Study and Gene Expression Analysis Identifies *CD84* as a Predictor of Response to Etanercept Therapy in Rheumatoid Arthritis

Jing Cui[1☯], Eli A. Stahl[1,2,3☯¤a], Saedis Saevarsdottir[4,5], Corinne Miceli[6,7], Dorothee Diogo[1,2,3], Gosia Trynka[1,2,3], Towfique Raj[2,3,8], Maša Umićević Mirkov[9], Helena Canhao[1,10,11], Katsunori Ikari[12], Chikashi Terao[13,14], Yukinori Okada[1,2,3], Sara Wedrén[4,5], Johan Askling[4,15], Hisashi Yamanaka[12], Shigeki Momohara[12], Atsuo Taniguchi[12], Koichiro Ohmura[13], Fumihiko Matsuda[14], Tsuneyo Mimori[13], Namrata Gupta[3], Manik Kuchroo[3,8], Ann W. Morgan[16], John D. Isaacs[17], Anthony G. Wilson[18], Kimme L. Hyrich[19], Marieke Herenius[20], Marieke E. Doorenspleet[20], Paul-Peter Tak[20¤b], J. Bart A. Crusius[21], Irene E. van der Horst-Bruinsma[22], Gert Jan Wolbink[23,24,25], Piet L. C. M. van Riel[9], Mart van de Laar[26], Henk-Jan Guchelaar[27], Nancy A. Shadick[1], Cornelia F. Allaart[28], Tom W. J. Huizinga[28], Rene E. M. Toes[28], Robert P. Kimberly[29], S. Louis Bridges Jr.[29], Lindsey A. Criswell[30], Larry W. Moreland[31], João Eurico Fonseca[10,11], Niek de Vries[20], Barbara E. Stranger[2,3], Philip L. De Jager[2,3,7], Soumya Raychaudhuri[1,2,3,32], Michael E. Weinblatt[1], Peter K. Gregersen[33], Xavier Mariette[6,7], Anne Barton[34], Leonid Padyukov[5], Marieke J. H. Coenen[9], Elizabeth W. Karlson[1], Robert M. Plenge[1,2,3]*

1 Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **3** Medical and Population Genetics Program, Chemical Biology Program, Broad Institute, Cambridge, Massachusetts, United States of America, **4** Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden, **5** Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, **6** Université Paris-Sud, Orsay, France, **7** APHP–Hôpital Bicêtre, INSERM U1012, Le Kremlin Bicêtre, Paris, France, **8** Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **9** Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, **10** Rheumatology Research Unit, Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisbon, Portugal, **11** Rheumatology Department, Santa Maria Hospital–CHLN, Lisbon, Portugal, **12** Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan, **13** Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan, **14** Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **15** Clinical Epidemiology Unit, Department of Medicine, Karolinska Institute/Karolinska University Hospital, Stockholm, Sweden, **16** NIHR–Leeds Musculoskeletal Biomedical Research Unit and Leeds Institute of Molecular Medicine, University of Leeds, Leeds, United Kingdom, **17** Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle Upon Tyne, United Kingdom, **18** Rheumatology Unit, Medical School, University of Sheffield, Sheffield, United Kingdom, **19** School of Translational Medicine, Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester, United Kingdom, **20** Department of Clinical Immunology and Rheumatology, Academic Medical Center/University of Amsterdam, Amsterdam, The Netherlands, **21** Laboratory of Immunogenetics, Department of Pathology, Vrije Universiteit Medical Center, Amsterdam, The Netherlands, **22** Department of Rheumatology, Vrije Universiteit University Medical Center, Amsterdam, The Netherlands, **23** Sanquin Research Landsteiner Laboratory, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, **24** School of Medicine and Biomedical Sciences, Sheffield University, Sheffield, United Kingdom, **25** Jan van Breemen Institute, Amsterdam, The Netherlands, **26** Arthritis Center Twente, University Twente and Medisch Spectrum Twente, Enschede, The Netherlands, **27** Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, The Netherlands, **28** Department of Rheumatology, Leiden University Medical Centre, Leiden, The Netherlands, **29** Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States of America, **30** Rosalind Russell Medical Research Center for Arthritis, Division of Rheumatology, Department of Medicine, University of California San Francisco, San Francisco, California, United States of America, **31** Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, **32** NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, United Kingdom, **33** The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, United States of America, **34** Arthritis Research UK Epidemiology Unit, Musculoskeletal Research Group, University of Manchester, Manchester Academic Health Sciences Centre, Manchester, United Kingdom

Abstract

Anti-tumor necrosis factor alpha (anti-TNF) biologic therapy is a widely used treatment for rheumatoid arthritis (RA). It is unknown why some RA patients fail to respond adequately to anti-TNF therapy, which limits the development of clinical biomarkers to predict response or new drugs to target refractory cases. To understand the biological basis of response to anti-TNF therapy, we conducted a genome-wide association study (GWAS) meta-analysis of more than 2 million common variants in 2,706 RA patients from 13 different collections. Patients were treated with one of three anti-TNF medications: etanercept (n = 733), infliximab (n = 894), or adalimumab (n = 1,071). We identified a SNP (rs6427528) at the *1q23* locus that was associated with change in disease activity score (ΔDAS) in the etanercept subset of patients ($P = 8 \times 10^{-8}$), but not in the infliximab or adalimumab subsets ($P > 0.05$). The SNP is predicted to disrupt transcription factor binding site motifs in the 3′ UTR of an immune-related gene, *CD84*, and the allele associated with better response to etanercept was associated with higher *CD84* gene expression in peripheral blood mononuclear cells ($P = 1 \times 10^{-11}$ in 228 non-RA patients and $P = 0.004$ in 132 RA patients). Consistent with the genetic findings, higher *CD84* gene expression correlated with lower cross-sectional DAS ($P = 0.02$, n = 210) and showed a non-significant trend for better ΔDAS in a subset of RA patients with gene expression data (n = 31, etanercept-treated). A small, multi-ethnic replication showed a non-significant trend towards an association among etanercept-treated RA

patients of Portuguese ancestry (n = 139, $P = 0.4$), but no association among patients of Japanese ancestry (n = 151, $P = 0.8$). Our study demonstrates that an allele associated with response to etanercept therapy is also associated with *CD84* gene expression, and further that *CD84* expression correlates with disease activity. These findings support a model in which *CD84* genotypes and/or expression may serve as a useful biomarker for response to etanercept treatment in RA patients of European ancestry.

Citation: Cui J, Stahl EA, Saevarsdottir S, Miceli C, Diogo D, et al. (2013) Genome-Wide Association Study and Gene Expression Analysis Identifies *CD84* as a Predictor of Response to Etanercept Therapy in Rheumatoid Arthritis. PLoS Genet 9(3): e1003394. doi:10.1371/journal.pgen.1003394

Editor: Alison Motsinger-Reif, North Carolina State University, United States of America

Received August 13, 2012; **Accepted** January 13, 2013; **Published** March 28, 2013

Funding: JC is supported by grants from ACR REF HPNIA award and the NIH (R01-AR059073, P60-AR047782, U01-GM092691, R01-AR049880). RMP is supported by grants from the NIH (R01-AR057108, R01-AR056768, U01-GM092691, R01-AR059648) and holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund. The American College of Rheumatology Research and Education Foundation provided funding to support this project. The eRA study was supported by R01 AI/AR47487. SS's work was supported by a clinical research fund from Stockholm County (ALF fund). LP's work was supported by the Swedish Rheumatism Association and the Swedish Medical Research Council. NdV and MED were sponsored by CTMM, the Center for Translational Molecular Medicine, and the Dutch Arthritis Foundation project TRACER (grant 04I-202). The Swedish studies (EIRA, Karolinska) were supported by grants from the Swedish Medical Research Council, the Stockholm County Council, the Swedish Council for Working Life and Social Research, King Gustaf V's 80-year foundation, the Swedish Rheumatism Association, the Swedish Foundation for Strategic Research, the Swedish COMBINE project, and the IMI funded BTCure project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rplenge@partners.org

¤a Current address: Division of Psychiatric Genomics, Mt. Sinai School of Medicine, New York, New York, United States of America
¤b Current address: GlaxoSmithKline, Stevenage, United Kingdom

☯ These authors contributed equally to this work.

Introduction

Rheumatoid arthritis (RA) is an autoimmune disease characterized by chronic inflammation of the synovial lining of the joint [1]. If left untreated, outcome varies from self-limited disease in a small proportion of RA patients to severe disease resulting in profound structural damage, excess morbidity and disability, and early mortality [2]. In the last twenty years, disease activity has been controlled in many patients by treatment with disease-modifying anti-rheumatic drugs (DMARDs), such as methotrexate, and the more recently developed biologic DMARDs that block inflammatory cytokines such as tumor necrosis factor-alpha (TNFa) [3]. Unfortunately, these medications are not effective in all RA patients, with up to one-third of patients failing to respond to any single DMARD [1–3]. Moreover, the biological mechanisms underlying treatment failure are unknown, which limits the development of clinical biomarkers to guide DMARD therapy or the development of new drugs to target refractory cases.

There are two classes of anti-TNF therapy: the TNF receptor fusion protein (etanercept), which acts as a soluble receptor to bind circulating cytokine and prevent TNF from binding to its cell surface receptor, and monoclonal antibodies that bind TNF (adalimumab, infliximab, certolizumab, and golimumab). There are undoubtedly shared mechanisms between the two drug classes (e.g., downstream signaling factors), as illustrated by similar effects on the change in inflammatory cytokines, complement activation, lymphocyte trafficking, and apoptosis [4,5,6]. Similarly, there are likely to be different biological factors that influence response: infliximab and adalimumab are approved for treatment of Crohn's disease; infliximab and adalimumab bind to transmembrane TNF on the surface of activated immune cells, whereas etanercept only binds soluble TNF [7]; and etanercept also binds a related molecule, lymphotoxin alpha (LTA), whereas infliximab/adalimumab do not [8].

Pharmacogenetics of response to anti-TNF therapy in RA remains in its early stages, with no single variant reaching an unambiguous level of statistical significance. Candidate gene studies suggest associations of TNFa or TNF receptor alleles, RA risk alleles or other SNPs with response to anti-TNF therapy [9,10,11]. Two GWAS in small sample sets (largest was 566 patients) have been performed, which identified loci with suggestive evidence for association [12,13]. Therefore, GWAS of large sample sizes may yet uncover genetic factors associated with response to anti-TNF therapy in RA, and larger cohorts enable separate analyses of the different types of anti-TNF drugs.

Here we report a GWAS of 2,706 samples with anti-TNF treatment response data collected from an international collaboration, including previously published GWAS data [12,13]. Our primary outcome measure was the change in disease activity score based on a joint count in 28 joints (DAS28) from baseline to 3–12 months after initiating anti-TNF therapy. Our secondary outcome measure was European League Against Rheumatism (EULAR) responder status [14,15], where patients are classified as EULAR good responders, moderate responders or non-responders based on follow up DAS28 after treatment and overall change in DAS28. We found a highly significant association for a variant that we also show is also a strong expression quantitative trait locus (eQTL) for the *CD84* gene. Our findings suggest that *CD84* genotype and/or expression may prove to be a biomarker for etanercept response in RA patients.

Results

Genome-wide association study

Clinical and GWAS data were compiled for 2,706 individuals of European ancestry from 13 collections as part of an international collaboration. Table 1 shows sample sizes, phenotypes and clinical variables for the four collections that were the units of analysis (additional details are shown in Table S1). Disease activity score based on a 28-joint count (DAS28) were collected at baseline and at one time point after anti-TNF therapy administration (mean 3.7 months, range 3–12 months). We defined our primary phenotype

Author Summary

There are no genetic predictors of response to one of the most widely used classes of drugs in the treatment of rheumatoid arthritis—biological modifiers of the inflammatory cytokine tumor necrosis factor-alpha (or anti-TNF therapy). To identify genetic predictors, we performed the largest genome-wide association study (GWAS) to date as part of an international collaboration. In our study, which included 2,706 RA patients treated with one of three anti-TNF drugs, the most significant finding was restricted to RA patients treated with etanercept ($P=8\times10^{-8}$), a drug that acts as a soluble receptor to bind circulating cytokine and prevents TNF from binding to its cell surface receptor. The associated variant influences expression of a nearby immune-related gene, *CD84*, whose expression is correlated with disease activity in RA patients. Together, our data support a model in which genomic factors related to *CD84* expression serve as a predictor of disease activity and response to etanercept therapy among RA patients of European ancestry, but not anti-TNF therapies that act through different biological mechanisms or potentially in RA patients of other genetic ancestries.

as a change in DAS28 (ΔDAS) from baseline (so that greater ΔDAS corresponded with better response to therapy; overall mean and standard deviation of 2.1 ± 1.3), adjusted for baseline DAS. A secondary phenotype was used based on European League Against Rheumatism (EULAR) response criteria. EULAR 'good response' was defined as ending DAS<3.2 and ΔDAS>1.2; 'non-response' was defined as ΔDAS <0.6 or ΔDAS≤1.2, and ending DAS >5.1; and 'moderate response' is in between [15]. We limited our secondary analysis to a dichotomous outcome, EULAR good responders (n = 998 for all patients treated with anti-TNF therapy) versus EULAR non-responders (n = 655), excluding the moderate category based on the hypothesis that a more extreme phenotype of response would yield improved discrimination.

Clinical variables were examined for association with phenotype, and therefore possible confounding in genetic association tests. In multivariate models (Table S2), only baseline DAS was strongly associated with the ΔDAS phenotype. As previously shown [11], age and gender showed univariate associations that were attenuated in the multivariate analysis. Accordingly, we used only baseline DAS as a clinical covariate, as this allowed us to maximize sample size given clinical variable missing data in some cohorts.

We performed quality control (QC) filtering and data processing of GWAS data for each of eleven genotyping batches. Genotyping array platforms are described in the Methods. HapMap2 imputation allowed us to test for association at >2 M SNPs with imputation quality scores >0.5. Genotype data were merged across several genotype batches to create four collections for genome-wide association testing. We performed linear regression association tests using baseline DAS and three principal components as covariates, and performed inverse-variance weighted meta-analysis to combine results across the four collections. Quantile-quantile plots with genomic control λ_{GC} values are shown in Figure S1. We found no evidence of systematic inflation of association test results, and no evidence of deflation for imputed versus genotyped SNPs. As a final filter, we excluded SNPs that

Table 1. Samples and clinical data.

Collection (analysis batch):	REF	BRAGGSS	DREAM	ReAct	Total
Sample sizes	959*	595	880*	272	2706
Drug subsets					
etanercept	365	259	109	0	733
infliximab	415	268	211	0	894
adalimumab	174	68	557	272	1071
EULAR Reponse categories					
Good responder	432**	161	313	92	998
Moderate responder	243	258	359	131	991
Non-responder	322	176	208	49	755
Genotype platform	mixed	Affy 500K	Illu550K +650K	Illumina OmniExpress	
Clinical variables					
Age, yr; mean (SD)	53.6 (12.7)	57.4 (10.9)	54.8 (12.9)	53.9 (10.8)	
Disease duration, yr; mean (SD)	6.7 (9.4)	14 (9.8)	9.6 (9.5)	12 (9.1)	
Gender, female %	75.6	77.3	68.3	77.9	
Seropositive, %	87	78	80	70	
MTX co-therapy, %	65.6	85.6	76.0	50.0	
Baseline DAS, mean (SD)	5.5 (1.2)	6.7 (0.9)	5.5 (1.2)	5.9 (1.0)	
ΔDAS, mean (SD)	1.9 (1.6)	2.5 (1.5)	1.9 (1.3)	2.2 (1.3)	
Mean treatment duration	4.6	5.6	3	3	
Study design	All***	Observational	Observational	Observational	

*8 patients had no TNF drug information.
**38 patients had only EULAR response (good, moderate or none) clinical data.
***ABCoN, GENRA are prospective cohorts, BeSt, eRA and TEAR are randomized controlled trial (RCT), and rest of REF group are observational cohorts.
doi:10.1371/journal.pgen.1003394.t001

showed strong evidence of heterogeneity across collections (Cochran's Q $P<0.001$).

We first analyzed all samples together (n = 2,706), regardless of drug type. We found no clear evidence of association with treatment response measured by ΔDAS (Figure 1A). Similar results were obtained using the binary phenotype of EULAR responder versus EULAR non-responder status (Figures S1 and S2).

We next separately analyzed patients treated with either etanercept (n = 733), infliximab (n = 894) or adalimumab (n = 1,071) (Figure 1B–1D), under the hypothesis that different genetic loci affect response to the different drugs based on their mechanism of action or other biochemical properties. GWAS results are publicly available for all SNPs tested at the Plenge laboratory and RICOPILI Web sites (see URLs). GWAS results for all SNPs achieving $P<10^{-6}$ from any analysis are detailed in the Table S3.

For etanercept-treated RA patients, a locus on chromosome *1q23* achieved near-genome-wide significance (rs6427528, $P_{META}=8\times10^{-8}$) (Figure 1B, Figure 2A, and Figure 3), but not in the infliximab or adalimumab subsets ($P>0.05$) (Figure S3). SNPs in linkage disequilibrium (LD) showed consistent association results (rs1503860, $P=1\times10^{-7}$, $r^2=1$ with rs6427528 in HapMap; three perfect-LD clusters of SNPs exemplified by rs3737792, rs10908787 and rs11265432 respectively; $P<5\times10^{-6}$; $r^2=0.83$, 0.63 and 0.59 with rs6427528, respectively). No single collection was responsible for the signal of association, as the effect size was consistent across all collections (Figure S4). The top SNP rs6427528 was genotyped in the ReAct dataset (Illumina Omni Express genotyping chip), and was well imputed across all other datasets (imputation quality score INFO ≥0.94, which is an estimate of genotype accuracy; the range of INFO scores is 0–1, where 1 indicates high confidence). All of these SNPs had minor

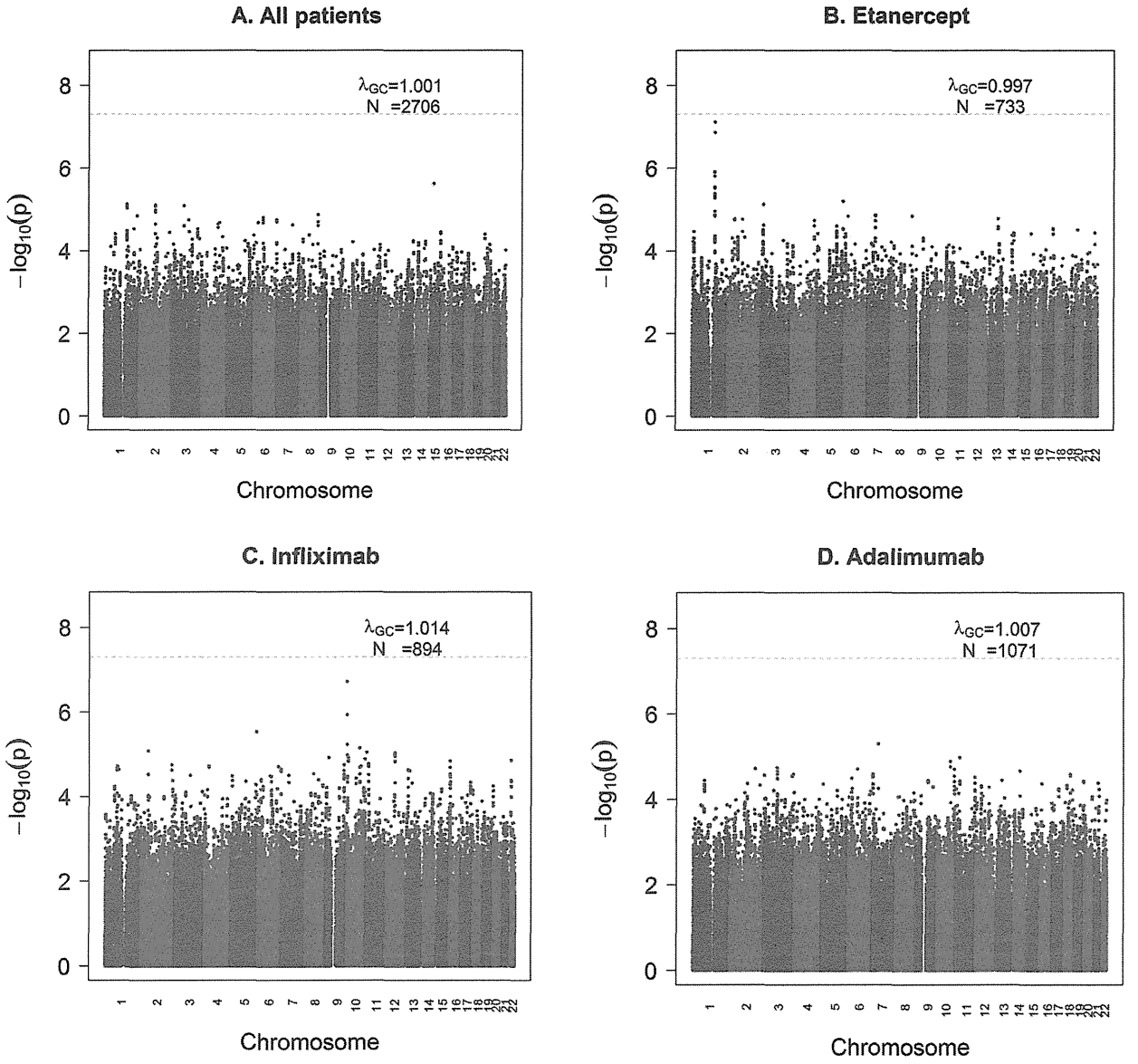
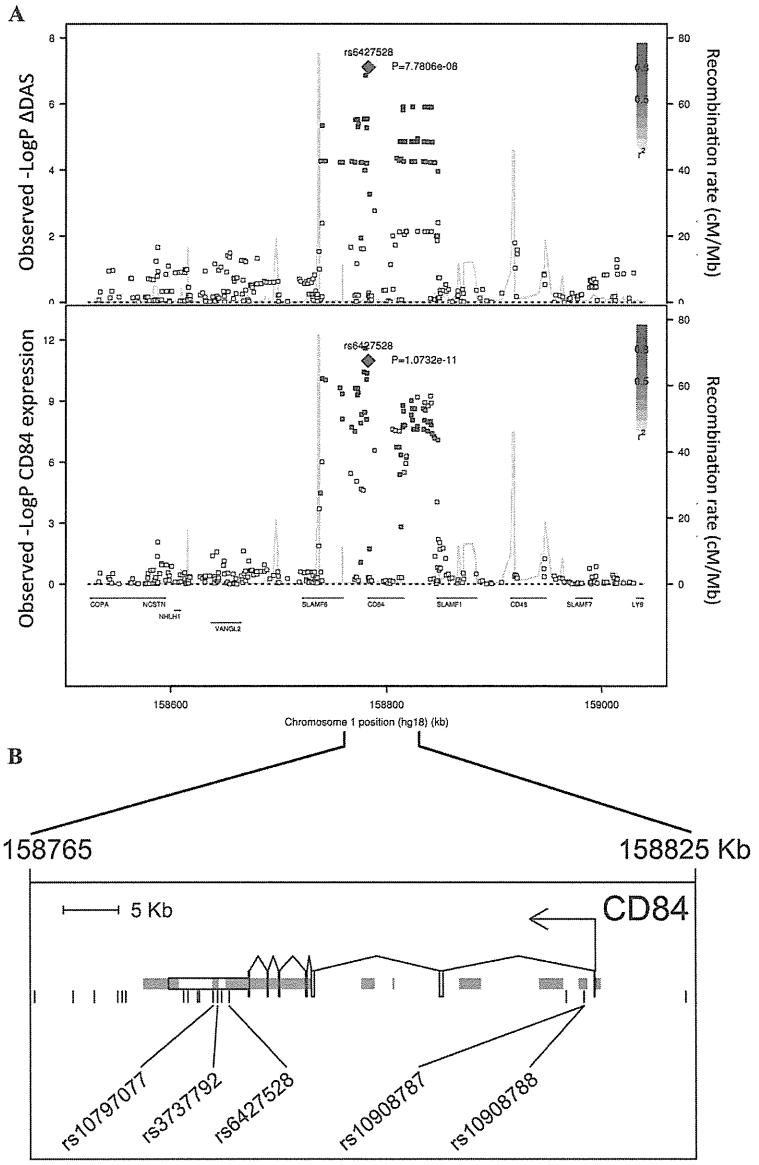


Figure 1. GWAS results for the ΔDAS phenotype. Shown are strengths of association (−Log10 P-value) for each SNP versus position along chromosomes 1 to 22. A) All samples (n = 2,706). B) Etanercept-treated patients (n = 733). C) Infliximab-treated patients (n = 894). D) Adalimumab-treated patients (n = 1,071).
doi:10.1371/journal.pgen.1003394.g001



C.

SNP (Major/ minor alleles)	Conserv-ation score[1]	DNase[2]	Transcription factor motifs altered		
			TF Motif	LOD(min) - LOD(maj)	Position weight matrix logo[3]
rs10797077 (G/A)	2.1	T-47D	AIRE_2	>6.6	
rs3737792 (G/A)	-1.3	-	-		
rs6427528 (A/G)	-2.8	-	KROX	>3	
			SREBP_4	-2.2	
rs10908787 (A/G)	-3.4	GM12878, Jurkat	-		
rs10908788 (A/G)	-7.8	GM12878	-		

Figure 2. Association results and SNP annotations in the *1q23 CD84* locus. A) Regional association plots with ΔDAS (top panel) and with *CD84* expression (bottom panel), showing strengths of association (−Log10 P-value) versus position (Kb) along chromosome 1. B) Schematic of *CD84* gene structure (RefSeq gene model, box exons connected by diagonal lines, arrow indicates direction of transcription) with strong enhancer chromatin states (orange rectangles) and SNPs in high LD (r2>0.8) with rs6427528 (vertical ticks). SNPs in enhancers are labeled below. C) Annotations of strong-enhancer rs6427528 proxy SNPs; listed are SNP rs-ID (major and minor alleles), conservation score, cell line with DNAse footprint if present, and transcription factor binding sites altered. 1- Genomic evolutionary rate profiling (GERP) conservation score, where a score >2 indicates conservation across mammals. 2- DNase footprint data are compiled from publicly available experiments by HaploReg. 3- Position weight matrix logos show transcription factor consensus binding sites with nucleotide bases proportional to binding importance. SNP position is boxed. Note that the rs10797077 AIRE_2 and the rs6427528 SREBP_4 motifs are on the minus strand (base complements correspond to SNP alleles), with the SREBP motif shown upside down to align with the rs6427528 KROX motif on the positive strand. Data are from HaploReg.
doi:10.1371/journal.pgen.1003394.g002

allele frequencies ranging from 7–10%. The SNP explains 2.6% variance in response to etanercept treatment.

For patients treated with infliximab, we observed a suggestive result on chromosome *10p14* (rs12570744, $P = 2\times10^{-7}$). No highly significant or suggestive results were observed for the ΔDAS phenotype in patients treated with adalimumab ($P_{META} > 10^{-5}$).

Qualitatively similar results were attained in the analysis of our secondary phenotype, EULAR good responder vs non-responder status (Figures S1 and S2). For SNPs at the *1q23* locus, the pattern of association with responder/non-responder status (etanercept-treated patients) was consistent with the results for ΔDAS ($P = 6\times10^{-3}$ for rs6427528 and rs1503860). We also identified potential novel associations, with suggestive results for infliximab (rs4336372, chromosome *5q35*, $P = 8\times10^{-7}$) and adalimumab (rs940928, chromosome *2q12*, $P = 2\times10^{-6}$).

eQTL and sequence analysis of the *CD84* gene

For each SNP with $P<10^{-6}$ identified by our GWAS (n = 6 independent SNPs), we searched for biological evidence to support a true positive association. We used genome-wide sequence data from the 1000 Genomes Project to search for putative functional variants in LD with the index SNP (defined as SNPs predicted to change protein-sequence or mRNA splicing). We also used genome-wide expression data to search for an expression quantitative trait locus (eQTL) in public databases and in peripheral blood mononuclear cells (PBMCs) in 228 non-RA patients and in 132 RA patients.

While we did not identify any variants disrupting protein-coding sequences or mRNA splicing, we did find that the *1q23* SNP associated with response to etanercept therapy was a strong eQTL in PBMCs (Figure 2A and Figure 3). In an analysis of 679 SNPs for cis-regulated expression of five genes in the region of LD (*SLAMF6*, *CD84*, *SLAMF1*, *CD48*, and *SLAMF7*), we found that rs6427528-*CD84* (and SNPs in LD with it) was the top eQTL of all results (n = 228 subjects; Figure 2A). This SNP was specifically associated with *CD84* expression, and was not an eQTL for other genes in the region (P>0.36 for the other genes).

We replicated our eQTL finding in 132 RA patients with both GWAS data and genome-wide expression data. PBMC expression data were available from RA patients in the Brigham RA Sequential Study (BRASS) and Autoimmune Biomarkers Collaborative Network (ABCoN) collections. We observed a significant association between rs6427528 genotype and *CD84* expression (linear regression adjusted for cohort P = 0.004, rank correlation P = 0.018). The direction of effect was the same as in the PBMC samples from 228 non-RA patients. A combined analysis of RA patients and the non-RA patient eQTL data (described above) yielded rank correlation $P = 3\times10^{-10}$ (n = 360 total individuals).

We searched sequence data to determine if rs6427528, or any of the SNPs in LD with it, were located within conserved, non-coding motifs that might explain the eQTL data. We used HaploReg [16] to examine the chromatin context of rs6427528 and 26 SNPs in LD with it (at $r^2>0.50$). We found that 5 SNPs occur in strong enhancers inferred from chromatin marks (Figure 2B) [17]. Two of these 5 SNPs, rs10797077 and rs6427528 ($r^2 = 0.74$ to each other), are predicted to disrupt transcription factor binding sites, and rs10797077 occurs at a site that shows conservation across mammalian genomes [18]. Figure 2C shows the DNA sequence position weight matrices of the transcription factor binding sites changed by rs10797077 (the minor allele creates a stronger binding site for the AIRE transcription factor) and rs6427528 (the minor allele creates a binding site for KROX and SREBP).

Expression of *CD84* as a biomarker of disease activity and treatment response

Because the genetic data demonstrates that the allele associated with better response is associated with higher *CD84* expression, this suggests that *CD84* expression itself may serve as a useful biomarker of disease activity or treatment response. We tested both hypotheses using PBMC expression data from the BRASS and ABCoN collections. First, we tested if *CD84* expression is associated with cross-sectional DAS, adjusting for age, gender and cohort (Figure 4). We observed a significant inverse association between *CD84* expression and cross-sectional DAS in 210 RA patients (beta = −0.3, P = 0.02, $r^2 = 0.02$). That is, higher *CD84* expression was associated with lower DAS, regardless of treatment.

Second, we tested *CD84* for association with our primary treatment response phenotype, ΔDAS. The sample size for this analysis was smaller than for the cross-sectional analysis, as we required that patients be on anti-TNF therapy and have pre- and post-treatment DAS. We found that *CD84* expression levels showed a non-significant trend towards an association with ΔDAS in 31 etanercept-treated patients (beta = 0.2, $r^2 = 0.002$, P = 0.46) and in all 78 anti-TNF-treated patients (beta = 0.14, $r^2 = 0.004$, P = 0.4). The effect is in the same direction one would predict based on the genetic association at rs6427528: the allele associated with better response is also associated with higher *CD84* expression (Figure 3), and in 31 RA patients, higher *CD84* expression (regardless of genotype) is associated with a larger ΔDAS (i.e., better response; Figure 4).

Replication of genetic data in a small, multi-ethnic cohort

Since most of the samples available to us as part of our international collaboration were included in our GWAS, few additional samples were available for replication. In addition, the remaining samples available to us were from different ethnic backgrounds. Nonetheless, we sought to replicate the associations of rs6427528 with ΔDAS in these additional samples. We genotyped 139 etanercept-treated patients from a rheumatoid arthritis registry in Portugal (Reuma.pt) and 151 etanercept-treated patients from two Japanese collections (IORRA, n = 88 patients on etanercept and Kyoto University, n = 63 on etanercept). Replication sample sizes, clinical data and results for these

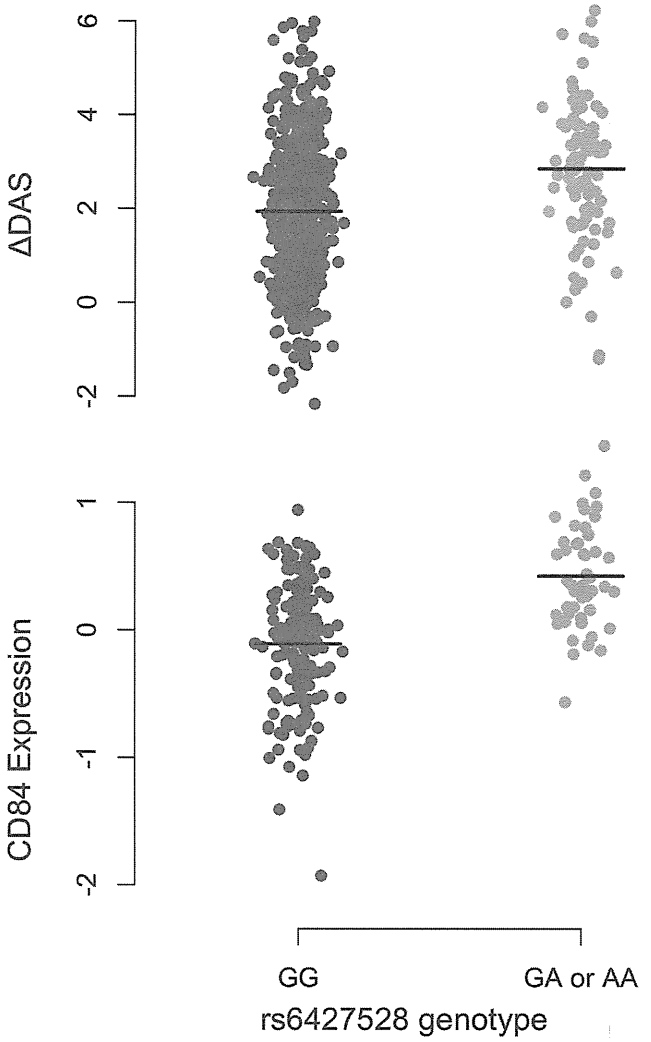


Figure 3. *1q23/CD84* genotype association plots for ΔDAS and *CD84* gene expression. Shown are ΔDAS in our GWAS in etanercept-treated patients (top panel, n = 733; n = 634 with the GG genotype and n = 99 with the GA or AA genotype) and *CD84* expression in our eQTL results (bottom panel, n = 228 non-RA patients; n = 178 with the GG genotype and n = 50 with the GA or AA genotype). The rare-allele homozygous genotype AA was observed four times in our ΔDAS GWAS and was pooled with the heterozygous GA genotype for this figure; AA homozygotes were not observed in the *CD84* eQTL data. Association analyses reported in the text regressed phenotype (ΔDAS, $P = 8 \times 10^{-8}$; *CD84* expression, $P = 1 \times 10^{-11}$) on minor-allele dosage (range 0–2).
doi:10.1371/journal.pgen.1003394.g003

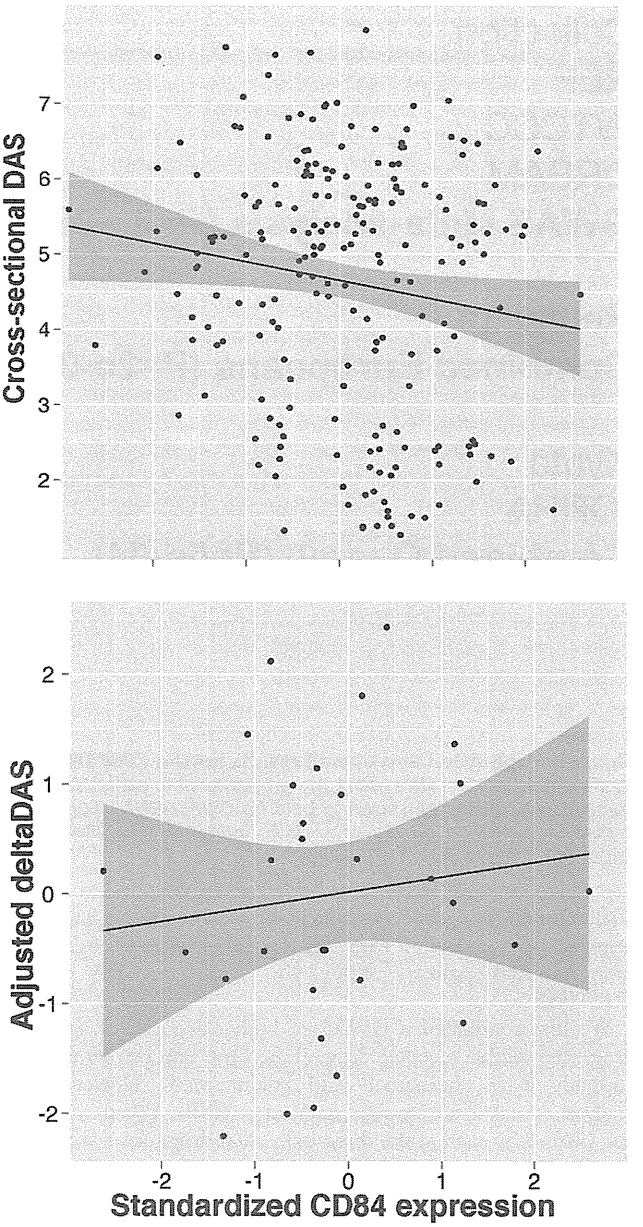


Figure 4. *CD84* expression level and clinical features. Analyses are shown in RA patients from the BRASS and ABCoN registries, for baseline DAS (top panel, n = 210; $R^2 = 0.02$, p = 0.02) and ΔDAS (bottom panel, n = 31; $R^2 = 0.001$, p = 0.46). Best-fit linear regression lines are shown in black, with shaded regions showing linear regression model (slope and intercept) 95% confidence intervals. *CD84* expression levels were quantile normalized, and ΔDAS values were adjusted for age, gender and baseline DAS.
doi:10.1371/journal.pgen.1003394.g004

two SNPs are shown in Table S4. Based on the observed effect size in the GWAS and observed allele frequency in the replication samples, we had 32% power to replicate this finding in the Portuguese samples and 17% power to replicate this finding in the Asian samples at $P<0.05$. The same association analysis as for GWAS was carried out: linear regression assuming an additive genetic model and using ΔDAS as phenotype, adjusted for baseline DAS. Replication results are shown in Figure 5.

While the SNPs fail to replicate in these patient collections at $P<0.05$, the direction of effect is the same in the Portuguese and Kyoto replication samples as in our GWAS. In a combined analysis limited to subjects of European-ancestry (GWAS data and Portuguese replication samples), rs6427528 remained highly suggestive ($P = 2 \times 10^{-6}$). Including the Japanese subjects, the overall GWAS+replication combined meta-analysis P-value remained suggestive ($P = 5 \times 10^{-4}$).

Discussion

Here we present the largest GWAS to date on anti-TNF therapy response in 2,706 RA patients. We find a significant association at the *1q23/CD84* locus in 733 etanercept treated patients ($P = 8 \times 10^{-8}$), but not in RA patients treated with drugs that act as a monoclonal antibody to neutralize TNF (infliximab or

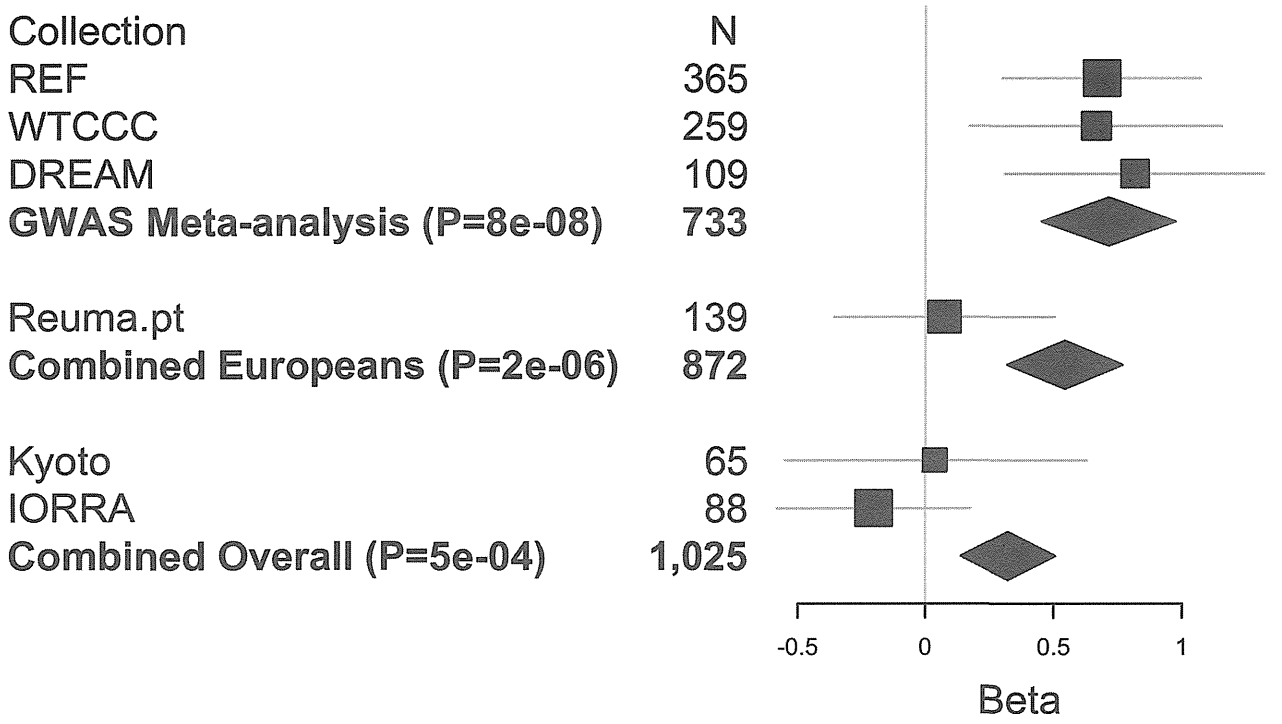


Figure 5. Replication and overall results for the *CD84* SNP rs6427528. Forest plot shows each cohort, sample size and linear regression beta coefficient estimates with symbol size proportional to cohort sample size and thin horizontal lines showing beta 95% CIs. Inverse variance weighted meta-analysis results are shown in bold for GWAS, GWAS+European (Portuguese) replication samples, and for GWAS+European+Asian (Japanese) replication samples, with diamond widths indicating beta 95% CIs.
doi:10.1371/journal.pgen.1003394.g005

adalimumab). The allele associated with a larger ΔDAS (i.e., better response) was associated with higher *CD84* expression in PBMCs from non-RA patients ($P = 1 \times 10^{-11}$) and in RA patients ($P = 0.004$).

We first conducted a GWAS of both categories of anti-TNF drugs (the soluble receptor drug, etanercept, and two monoclonal antibody drugs, infliximab and adalimumab). However, this analysis revealed no strongly associated SNPs. When we subset our GWAS by each of the three individual drugs, several SNPs in the *1q23* locus were highly significant in etanercept-treated patients, and SNPs in three other loci (*10p15*, *5q35* and *2q12*) were associated in infliximab or adalimumab subset analyses. Furthermore, the top SNPs for each analysis (Table S3) showed little correlation across the three anti-TNF drugs. This simple observation suggests that genetic control of treatment response may be different for different drugs. This finding is consistent with the clinical observation that RA patients who fail one anti-TNF drug may still respond to a different anti-TNF drug, albeit at lower rates of response [19]. If confirmed in larger samples and more comprehensive analyses, then this could have major implications for how physicians prescribe these drugs.

The most significant finding from our GWAS was a set of equivalent SNPs in LD with each other from the *1q23* locus in etanercept-treated RA patients (Figure 1 and Figure 2A). While the top SNP did not reach genome-wide significance in predicting treatment response, it did reach genome-wide significance as an eQTL in PBMCs ($P = 1 \times 10^{-11}$; Figure 2A). This finding indicates that the SNP (or another variant in LD with it) is indeed biologically functional in a human tissue that is important in the immune response. Two SNPs, rs10797077 and rs6427528, disrupt transcription factor binding sites, and represent excellent candidates for the causative allele to explain the effect on *CD84* expression (Figure 2C).

Our findings suggest that *CD84* genotype and/or expression could be a biomarker for etanercept treatment response among individuals of European ancestry. The genetic and expression data predict that *CD84* expression should be positively associated with treatment response (i.e., higher expression is associated with better response; Figure 3). While we did not observe a significant association between *CD84* levels and ΔDAS, we did observe a trend consistent with this prediction (Figure 4). Importantly, we note that power was extremely limited with the small sample sizes for which we had *CD84* expression as well as drug response data ($n = 31$ RA patients treated with etanercept).

The *CD84* gene is a compelling candidate for immune response, belonging to the CD2 subset of the immunoglobulin superfamily. It has been implicated in T-cell activation and maturation [20]. CD84 localizes to the surface of CD4+ and CD8+ T cells, and acts as a costimulatory molecule for IFN-gamma secretion [21]. *CD84* is also expressed in B-cells, monocytes and platelets. *CD84* has not been previously implicated in genetic studies of RA risk, disease activity, disease severity, or treatment response.

A limitation of our study is the small sample size available for replication ($n = 290$ etanercept-treated patients), and the lack of replication observed for the top *CD84* SNP (rs6427528) among patients of Portuguese and Japanese ancestry. The simplest explanation is that our original observation in the GWAS data represents a false positive association. However, the eQTL and gene expression data argue against this possibility. Explanations for a false negative finding in our replication collections include: (1) lack of power, especially if the effect size observed in the GWAS represents an over-estimate of the true effect size (the Winner's Curse) – we estimate that we had 32% and 17% power (at $P = 0.05$) to detect an association in the Portuguese and Japanese sample collections, respectively; (2) clinical heterogeneity, which is

always a possibility in pharmacogenetic studies, especially those conducted in different countries; and (3) ethnic differences, including different patterns of LD between the underlying causative allele (which is as yet unknown) and marker SNPs tested in our study. We did observe subtle differences in local patterns of LD between Asians and Europeans using genetic data from the 1000 Genomes Project (Figure S5). We note that the rs6427528 minor allele A has a frequency of ~5–10% in European and East Asian populations, and ~50% in the African YRI population (HapMap2 and 1000 Genomes); therefore, it may be of interest to test African American samples in replication.

What are the options for increasing sample size in pharmacogenetic studies, thereby providing an opportunity to replicate our *CD84* genetic and expression findings? While it might seem trivial to collect more samples through traditional registries, this is extremely challenging for phenotypes pertaining to treatment efficacy. To underscore this point, we highlight our study design, where we organized samples and clinical data from 16 different collections across 7 different countries in order to obtain the samples for the current study. Going forward, non-traditional strategies to collect biospecimens linked with clinical data (e.g., online registries, electronic medical records) may be required to achieve clinical collections of sufficient size to discover pharmacogenomic predictors of efficacy.

In conclusion, we conducted the largest GWAS to date for response to anti-TNF therapy in RA patients. Our genetic and expression data suggest that *CD84* genetic variants and/or expression levels could be developed as predictive biomarkers for etanercept treatment response in RA patients of European ancestry.

Methods

Samples and clinical data

All patients met 1987 ACR criteria for RA, or were diagnosed by a board-certified rheumatologist. In addition, patients were required to have at least moderate disease activity at baseline (DAS>3.2). All patients gave their informed consent and all institutional review boards approved of this study. A total of 13 collections from across 5 countries were included in GWAS [11,12,13,22]: Autoimmune Biomarkers Collaborative Network (ABCoN) from the U.S. (N = 79); the Genetics Network Rheumatology Amsterdam (GENRA, N = 53); the Dutch Behandelstrategieen voor Rheumatoide Arthritis (BeSt, N = 85); the U.K. Biological in Rheumatoid arthritis Genetics and Genomics Study Syndicate (BRAGGSS, N = 140); the U.S. Brigham Rheumatoid Arthritis Sequential Study (BRASS, N = 55); the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA, N = 298); the Immunex Early Rheumatoid Arthritis study (eRA N = 57); the Swedish Karolinska Institutet study (KI, N = 77); the Netherlands collection from Leiden University Medical Center (LUMC, N = 43); and the U.S. Treatment of Early Aggressive RA (TEAR, N = 109). We refer to these collections as the American College of Rheumatology Research and Education Foundation (REF) collection, as funding for GWAS genotyping was provided by the "*Within Our Reach*" project. We included additional samples from BRAGGSS (N = 595) [12]; the Dutch Rheumatoid Arthritis Monitoring registry (DREAM) in the Netherlands, and the ApotheekZorg (AZ) database (which facilitates the Dutch distribution of adalimumab; N = 880) [23,24], together referred to as DREAM; and the French Research in Active Rheumatoid Arthritis (ReAct, N = 272) [25].

Additional samples were collected for replication of SNPs in the *1q23* locus. These included the Rheumatic Diseases Portuguese Register (Reuma.pt, N = 378) from the Portuguese Society of Rheumatology (SPR), which captures more than 90% of patients treated with biological therapies and managed in rheumatology departments across Portugal [26]. Additional replication samples (N = 374) of East Asian ancestry were included from the IORRA and Kyoto University Hospital registries, part of the Japanese Genetics and Allied research in Rheumatic diseases Networking consortium (GARNET) [27].

Clinical data were collected in each cohort, including disease activity scores at baseline and at least one time point after treatment, gender, age, methotrexate use, as well as autoantibody status (RF or CCP). The composite disease activity scores for 28 joints (DAS28) included laboratory values for erythrocyte sedimentation rate (ESR) for most samples and C-reactive protein (CRP) for 191 samples in the REF collection (ABCoN, BRASS and eRA cohorts). DAS28 values were available at baseline and at 3–12 months after initiating anti-TNF therapy. Our primary phenotype was defined as ΔDAS = baseline DAS - end DAS, and responder status was also determined according to EULAR criteria for start and end DAS [15]. Clinical variables were assessed for association with phenotype in multivariate linear or logistic regression models for both the ΔDAS and EULAR responder-status phenotypes. Clinical variables that were significant in these analyses were retained as covariates in genetic association tests, except for methotrexate co-therapy. Including a covariate for methotrexate co-therapy reduced sample size substantially due to missing clinical data, so results were compared for our primary analysis and a secondary analysis with the covariates (and with reduced sample size) and the results were verified not to be impacted (not shown).

Genotyping and data processing

A total of eleven genotyping batches were processed separately. (1) BRASS samples were genotyped using Affymetrix 6.0 chip [28]; (2) WTCCC samples were genotyped on Affymetrix 500K chip [12]. All other cohorts were genotyped using Illumina platform arrays (see Table 1). Our American College of Rheumatology Research Education Fund (REF) collection was made up of smaller cohorts from throughout North America and Europe, including BRASS samples. Also included in REF: (3) ABCoN [13] and (4) EIRA [29] were separately genotyped on the Illumina 317K genotyping array; (5) eRA on the Illumina 550K chip; and (6) GENRA, BeSt, BRAGGSS (a subset of N = 53 samples), KI and LUMC were genotyped in one batch, and (7) BRAGGSS (N = 87) and TEAR were genotyped in a second batch, both using Illumina 660k chips, at the Broad Institute (8–10). DREAM and AZ samples were genotyped in three batches, one on 550K chip and two on 660K chips (manuscript in preparation), and (11) ReAct samples were genotyped on Illumina OmniExpress chips. Quality control (QC) filtering was done in each genotyping batch, including filtering individuals with >5% missing data, and filtering SNPs with >1% missing data, minor allele frequency (MAF) <1% and Chi-squared test of Hardy Weinberg equilibrium $P_{HWE}<10^{-5}$. We then used individual-pairwise identity-by-state estimates to remove occasional related and potentially contaminated samples. Data processing and QC were performed in PLINK [30]. Principal Components Analysis (PCA) was performed using EIGENSTRAT [31] (default settings) on the combined dataset using 20,411 SNPs genotyped across all datasets. Ethnicity outliers including all individuals of non-European decent were identified and removed, and the first three eigenvectors were used as covariates in GWAS.

Imputation was conducted on each of eleven datasets separately, using the IMPUTE v1 software [32] and haplotype-phased

HapMap Phase 2 (release 22) European CEU founders as a reference panel. Imputation of BRASS and EIRA was previously reported [28,33], and we followed the same imputation procedures for the remaining datasets. Imputation yielded posterior genotype probabilities as well as imputation quality scores at SNPs not genotyped with a minor allele frequency $\geq 1\%$ in HapMap CEU. We removed imputed SNPs with imputation 'info' scores <0.5 or MAF $<1\%$ in any of the datasets.

Expression profile and eQTL data

Gene expression levels were quantified using mRNA derived from peripheral blood mononuclear cells (PBMCs) using Affymetrix Human Genome U133 Plus 2.0, for 255 multiple sclerosis patients in the Comprehensive Longitudinal Investigation of MS at the Brigham and Women's Hospital [34], either untreated (N = 83) or treated with interferon-beta (N = 105) or glatiramer acetate (N = 67). The raw intensity values were subject to quality control based on the recommended pipeline available in the simpleaffy and affyPLM R Bioconductor packages, and were then normalized using GCRMA (N = 228). The data are available on the Gene Expression Omnibus website (GSE16214). Expression levels for 17,390 probes mapping to 9,665 Ensembl transcripts were adjusted for confounding factors including age, gender, drug and batch using principle components and Bayesian factor analysis [35], and used in eQTL association analyses. Genotype data were collected on the Affymetrix 550K GeneChip 6.0 platform as a part of a previously published study [36]. Allelic dosages from imputed data (HapMap Phase II CEU samples; >2 million SNPs, MACH imputation quality >0.1 and MAF> = 0.05) were used for association analysis. *Cis*-eQTLs were identified +/−1 Mb of transcription start sites (TSS) in the *1q23* locus region. Significance was evaluated by 10,000 permutations per gene, and false discovery rates were calculated based on cis-eQTL analyses in the total of 9,665 genes [37].

Additional expression profile data were available for subsets of samples that were part of two cohorts in our GWAS. Expression data from patients enrolled in the BRASS registry have been previously published [38]. Expression data were collected on Affymetrix Gene Chip U133 Plus 2 microarrays. BRASS patients had either cross-sectional expression data (n = 132, assayed at the time the patient was enrolled in BRASS) or pre- and post-treatment expression data (n = 17 samples, 8 treated with etanercept). Of these, n = 87 patients had expression and GWAS data. For patients with pre- and post-treatment data, we used the "baseline" pre-treatment expression data for cross-sectional analysis. In ABCoN, 65 RA patients (n = 23 treated with etanercept) had both pre- and post-treatment expression data, as well as ΔDAS clinical data [39], and n = 45 patients had expression and GWAS data. As with BRASS, we use the "baseline" pre-treatment expression data for cross-sectional analysis. For ABCoN expression profile data were collected on Illumina Human WG6v3 microarrays and were quantile normalized according to Illumina recommended protocols. Within both BRASS and ABCoN, expression data were normalized to the mean and standard deviation within each collection. For prospective analyses of expression data and ΔDAS, we combined BRASS and ABCoN to include 31 etanercept-treated patients and 78 anti-TNF-treated patients.

Statistical analyses

In our primary GWAS analysis, we tested each SNP for association with ΔDAS using linear regression adjusted for baseline DAS and the first 3 PCA eigenvectors in each collection. In our secondary GWAS analysis, we modeled SNPs predicting EULAR good response *versus* EULAR non-response using logistic regression, again adjusting for start-DAS value and the first three eigenvectors. Association analysis was done using SNPTEST [32] assuming an additive genetic model. Genomic control λ_{GC} values [40] for genotyped SNPs only and all SNPs were calculated, and no inflation or deflation was observed in the distributions of association test results. We then conducted inverse variance-weighted meta-analysis to combine results across the four datasets, and conducted Cochran's Q tests for heterogeneity using the β coefficients [41]. We further divided samples into 3 subsets according to drug (etanercept, infliximab or adalimumab). GWAS analysis for each group followed the same analysis procedure. Meta-analysis and heterogeneity tests were conducted using SAS. Expression analyses utilized linear regression or Spearman rank correlation, also using SAS. We tested for effects of cohort, age, gender and concurrent methotrexate, and results are shown using significant covariates as indicated.

Supporting Information

Figure S1 Quantile–quantile (QQ) plots for ΔDAS and response analysis, with genomic control λ_{GC} values.
(TIF)

Figure S2 GWAS results for the good response versus non-response phenotype. Shown are strengths of association (−Log10 P-value) for each SNP versus position along chromosomes 1 to 22. A) All samples (n = 1,708). B) Etanercept-treated patients (n = 472). C) Infliximab-treated patients (n = 599). D) Adalimumab-treated patients (n = 636).
(TIF)

Figure S3 Forest plot of replication results for the CD84 SNP rs6427528, in patients treated with anti-TNF drugs other than etanercept (infliximab & adalimumab).
(TIF)

Figure S4 Forest plot of CD84 result in patients treated with etanercept, subset by all collections.
(TIF)

Figure S5 Patterns of linkage disequilibrium (LD) at the CD84 locus in HapMap. Shown patterns of LD for CEU (top panel) and CHBJPT (bottom panel).
(TIF)

Table S1 Sample information for each of thirteen clinical batches.
(DOC)

Table S2 Clinical multivariate model for the ΔDAS phenotype.
(DOC)

Table S3 GWAS results for all SNPs achieving $P<10^{-6}$ from any analysis.
(XLS)

Table S4 Sample and clinical data summary for replication samples.
(DOC)

Acknowledgments

We thank Lars Klareskog and Lars Alfredsson for their contribution to the EIRA cohort, and the Swedish Rheumatology Register and Biologics Register (ARTIS) for providing follow-up data for the patients in the EIRA and KI study groups. We thank participants in the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate (BRAGGSS).

Author Contributions

Conceived and designed the experiments: JC EAS RMP. Performed the experiments: JC EAS RMP. Analyzed the data: JC EAS RMP. Contributed reagents/materials/analysis tools: SS CM DD GT TR MUM HC KI CT YO SW JA HY SM AT KO FM TM NG MK AWM JDI AGW KLH MH MED P-PT JBAC IEvdH-B GJW PLCMvR MvdL H-JG NAS CFA TWJH REMT RPK SLB LAC LWM JEF NdV BES PLDJ SR MEW PKG XM AB LP MJHC EWK. Wrote the paper: JC EAS RMP. All authors reviewed and approved the manuscript.

References

1. Klareskog L, Catrina AI, Paget S (2009) Rheumatoid arthritis. Lancet 373: 659–672.
2. Scott DL, Wolfe F, Huizinga TW (2010) Rheumatoid arthritis. Lancet 376: 1094–1108.
3. McInnes IB, Schett G (2011) The pathogenesis of rheumatoid arthritis. N Engl J Med 365: 2205–2219.
4. Aeberli D, Seitz M, Juni P, Villiger PM (2005) Increase of peripheral CXCR3 positive T lymphocytes upon treatment of RA patients with TNF-alpha inhibitors. Rheumatology (Oxford) 44: 172–175.
5. Agnholt J, Dahlerup JF, Kaltoft K (2003) The effect of etanercept and infliximab on the production of tumour necrosis factor alpha, interferon-gamma and GM-CSF in in vivo activated intestinal T lymphocyte cultures. Cytokine 23: 76–85.
6. Catrina AI, Trollmo C, af Klint E, Engstrom M, Lampa J, et al. (2005) Evidence that anti-tumor necrosis factor therapy with both etanercept and infliximab induces apoptosis in macrophages, but not lymphocytes, in rheumatoid arthritis joints: extended report. Arthritis Rheum 52: 61–72.
7. Scallon BJ, Moore MA, Trinh H, Knight DM, Ghrayeb J (1995) Chimeric anti-TNF-alpha monoclonal antibody cA2 binds recombinant transmembrane TNF-alpha and activates immune effector functions. Cytokine 7: 251–259.
8. Gudbrandsdottir S, Larsen R, Sorensen LK, Nielsen S, Hansen MB, et al. (2004) TNF and LT binding capacities in the plasma of arthritis patients: effect of etanercept treatment in juvenile idiopathic arthritis. Clin Exp Rheumatol 22: 118–124.
9. Plant D, Prajapati R, Hyrich KL, Morgan AW, Wilson AG, et al. (2012) Replication of association of the PTPRC gene with response to anti-tumor necrosis factor therapy in a large UK cohort. Arthritis Rheum 64: 665–670.
10. Prajapati R, Plant D, Barton A (2011) Genetic and genomic predictors of anti-TNF response. Pharmacogenomics 12: 1571–1585.
11. Cui J, Saevarsdottir S, Thomson B, Padyukov L, van der Helm-Van Mil AH, et al. (2010) Rheumatoid arthritis risk allele PTPRC is also associated with response to anti-tumor necrosis factor alpha therapy. Arthritis Rheum 62: 1849–1861.
12. Plant D, Bowes J, Potter C, Hyrich KL, Morgan AW, et al. (2011) Genome-wide association study of genetic predictors of anti-tumor necrosis factor treatment efficacy in rheumatoid arthritis identifies associations with polymorphisms at seven loci. Arthritis Rheum 63: 645–653.
13. Liu C, Batliwalla F, Li W, Lee A, Roubenoff R, et al. (2008) Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. Mol Med 14: 575–581.
14. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, et al. (1995) Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 38: 44–48.
15. van Gestel AM, Prevoo ML, van 't Hof MA, van Rijswijk MH, van de Putte LB, et al. (1996) Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum 39: 34–40.
16. Ward LD, Kellis M (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 40: D930–934.
17. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473: 43–49.
18. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, et al. (2005) Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 15: 901–913.
19. Soliman MM, Hyrich KL, Lunt M, Watson KD, Symmons DP, et al. (2012) Rituximab or a second anti-TNF therapy for rheumatoid arthritis patients who have failed their first anti-TNF? Comparative analysis from the British Society for Rheumatology Biologics Register. Arthritis Care Res (Hoboken).
20. Tangye SG, Nichols KE, Hare NJ, van de Weerdt BC (2003) Functional requirements for interactions between CD84 and Src homology 2 domain-containing proteins and their contribution to human T cell activation. J Immunol 171: 2485–2495.
21. Martin M, Romero X, de la Fuente MA, Tovar V, Zapater N, et al. (2001) CD84 functions as a homophilic adhesion molecule and enhances IFN-gamma secretion: adhesion is mediated by Ig-like domain 1. J Immunol 167: 3668–3676.
22. Padyukov L, Lampa J, Heimburger M, Ernestam S, Cederholm T, et al. (2003) Genetic markers for the efficacy of tumour necrosis factor blocking therapy in rheumatoid arthritis. Ann Rheum Dis 62: 526–529.
23. Coenen MJ, Enevold C, Barrera P, Schijvenaars MM, Toonen EJ, et al. (2010) Genetic variants in toll-like receptors are not associated with rheumatoid arthritis susceptibility or anti-tumour necrosis factor treatment outcome. PLoS ONE 5: e14326. doi:10.1371/journal.pone.0014326
24. Toonen EJ, Coenen MJ, Kievit W, Fransen J, Eijsbouts AM, et al. (2008) The tumour necrosis factor receptor superfamily member 1b 676T>G polymorphism in relation to response to infliximab and adalimumab treatment and disease severity in rheumatoid arthritis. Ann Rheum Dis 67: 1174–1177.
25. Miceli-Richard C, Comets E, Verstuyft C, Tamouza R, Loiseau P, et al. (2008) A single tumour necrosis factor haplotype influences the response to adalimumab in rheumatoid arthritis. Ann Rheum Dis 67: 478–484.
26. Canhao H, Faustino A, Martins F, Fonseca JE (2011) Reuma.pt - the rheumatic diseases portuguese register. Acta Reumatol Port 36: 45–56.
27. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, et al. (2012) Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet.
28. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, et al. (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 42: 508–514.
29. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, et al. (2007) TRAF1-C5 as a Risk Locus for Rheumatoid Arthritis – A Genomewide Study. N Engl J Med 357: 1199–1209.
30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
31. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909.
32. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39: 906–913.
33. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, et al. (2012) Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet.
34. Gauthier SA, Glanz BI, Mandel M, Weiner HL (2006) A model for the comprehensive investigation of a chronic autoimmune disease: the multiple sclerosis CLIMB study. Autoimmun Rev 5: 532–536.
35. Stegle O, Parts L, Durbin R, Winn J (2010) A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol 6: e1000770. doi:10.1371/journal.pcbi.1000770
36. De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, et al. (2009) Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet 41: 776–782.
37. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39: 1217–1224.
38. Parker A, Izmailova ES, Narang J, Badola S, Le T, et al. (2007) Peripheral Blood Expression of Nuclear Factor-kappaB-Regulated Genes Is Associated with Rheumatoid Arthritis Disease Activity and Responds Differentially to Anti-Tumor Necrosis Factor-alpha versus Methotrexate. J Rheumatol 34: 1817–1822.
39. Batliwalla FM, Baechler EC, Xiao X, Li W, Balasubramanian S, et al. (2005) Peripheral blood gene expression profiling in rheumatoid arthritis. Genes Immun 6: 388–397.
40. Devlin B, Roeder K, Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 60: 155–166.
41. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, et al. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 17: R122–128.