

Proceedings

Open Access

## Conditional analysis of the major histocompatibility complex in rheumatoid arthritis

Kimberly E Taylor and Lindsey A Criswell\*

Address: The Rosalind Russell Medical Research Center for Arthritis, University of California San Francisco, 374 Parnassus Ave, San Francisco, California 94143 USA

E-mail: Kimberly E Taylor - kim.taylor@ucsf.edu; Lindsey A Criswell\* - lindsey.criswell@ucsf.edu

\*Corresponding author

from Genetic Analysis Workshop 16  
St Louis, MO, USA 17-20 September 2009

Published: 15 December 2009

BMC Proceedings 2009, 3(Suppl 7):S36 doi: 10.1186/1753-6561-3-S7-S36

This article is available from: <http://www.biomedcentral.com/1753-6561/3/S7/S36>

© 2009 Taylor and Criswell; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

We performed a whole-genome association study of rheumatoid arthritis susceptibility using Illumina 550k single-nucleotide polymorphism (SNP) genotypes of 868 cases and 1194 controls from the North American Rheumatoid Arthritis Consortium (NARAC). Structured association analysis with adjustment for potential population stratification yielded 200 SNPs with  $p < 1 \times 10^{-8}$  for association with RA, all of which were on chromosome 6 in a 2.7-Mb region of the major histocompatibility complex (MHC). Given the extensive linkage equilibrium in the region and known risk of *HLA-DRB1* alleles, we then applied conditional analyses to ascertain independent signals for RA susceptibility among these 200 candidate SNPs. Conditional analyses incorporating risk categories of the *HLA-DRB1* "shared epitope" revealed three SNPs having independent associations with RA (conditional  $p < 0.001$ ). This supports the presence of significant effects on RA susceptibility in the MHC in addition to the shared epitope.

### Background

Rheumatoid arthritis (RA) is a chronic systemic autoimmune disease characterized by damage to synovial joints as well as extraarticular manifestations. The strongest known genetic risk factor is the *HLA-DRB1* gene on chromosome 6, namely a set of alleles sharing a common sequence known as the shared epitope (SE) [1]. Recent whole-genome association studies have revealed new risk genes outside of the HLA region [2-4], and some studies have also provided evidence of additional influences from the HLA class III and class I regions [5,6]. In this analysis we sought first to identify and/or

validate RA risk alleles throughout the genome, and then to identify independent associations with RA susceptibility in the major histocompatibility complex (MHC) in addition to the SE.

### Methods

Illumina 550k genotyping data from a whole-genome association study by North American Rheumatoid Arthritis Consortium (NARAC) [3] was used for this study as part of the Genetic Analysis Workshop 16; duplicated and contaminated samples had been removed previously. Using the computer program

PLINK [7], subjects were filtered who had less than 90% genotyping, and single-nucleotide polymorphisms (SNPs) were filtered that had less than 90% genotyping, Hardy-Weinberg equilibrium in controls  $p < 0.0001$ , or minor allele frequency (MAF)  $< 0.05$ . Using the computer program EIGENSTRAT [8], population outliers were removed who were  $>6$  standard deviations from any of the first five principal components (PCs) identified in PC analysis.

First we analyzed the whole-genome data using structured association analyses of EIGENSTRAT; although the NARAC cases and controls are Caucasian, differences in intra-European ancestry [9] can produce false-positive associations. SNPs used for the PCA were filtered to remove regions of extended local associations (chr 8: 8-12 Mb, chr 6: 24-36 Mb, chr 11: 42-58 Mb, chr 5: 44-51.5 Mb, and chr 17: 40-43 Mb) and pruned to have  $r^2 < 0.2$  within a sliding window of 1 kb with a step size of 100, similar to methods of Fellay et al. [10] and Hom et al. [11]. We included correction for the first six PCs (see Results).

We used a conservative genome-wide significance threshold of  $p = 1 \times 10^{-8}$ . Because all SNPs exceeding this threshold (i.e., lower  $p$ -value) were in the MHC in a region of extended linkage disequilibrium (LD), we proceeded with conditional analyses to attempt to establish signals that were independent of the shared epitope and each other. We modeled the SE as a multi-allelic marker with values corresponding to negative, low-risk, or high-risk. SE alleles were considered high risk if they were one of DRB1\*0401, 0404, 0405, 0408, or 0409. Table 1 shows the case-control ratios for each risk category.

Our conditional analyses, using the computer program Whap [12], proceeded as follows. Starting with the SE as the top marker, we tested each two-SNP marker (SE plus each other SNP) for independence of the other SNP; in particular this uses a likelihood-ratio test to determine the significance of the difference between the two-SNP "alternate" model versus the one-SNP (SE in this case) "null" model. As long as the most significant SNP was  $<0.001$ , we added this best SNP to the list of independent SNPs, and proceeded to test all three-

marker combinations compared to our best two-SNP model; and so on with larger haplotypes. In the final list, we also tested each locus for a significant addition to the model containing all other top SNPs.

### Results

After quality control filtering (above), 486,078 SNPs remained for the whole-genome analyses. All subjects had genotyping  $>90\%$ , and eight controls were removed as outliers detected by EIGENSTRAT, leaving 868 cases and 1186 controls for the final analyses. In structured association analysis with EIGENSTRAT, we corrected for the first six PCs because the scree graph of eigenvalues levelled off at the sixth component. PCs one, two, four, and five were all highly significant in association tests with cases and controls (all  $p \leq 10^{-8}$ ). In this whole-genome analysis we identified exactly 200 SNPs with  $p < 10^{-8}$ , all between 30.38 Mb and 33.08 Mb in the MHC region. Table 2 shows the significance of associations in this dataset for known RA risk alleles outside of the MHC [2,3,13-17]; for SNPs not in the Illumina 550k panel, there were perfect proxy SNPs ( $r^2 = 1$ ) in the HapMap CEU population [18]. Although not reaching our  $10^{-8}$  threshold, we observed  $p$ -values from  $10^{-5}$  to  $5 \times 10^{-6}$  for *PTPN22* and *TRAF1-C5* SNPs, and  $p = 0.03$  for *STAT4*. This dataset was underpowered to detect any of these risk alleles at a genome-wide level for their published odds ratios (ORs); for example, we had approximately 70% power to detect the highest OR of 1.75 (*PTPN22*, MAF = 11%) at  $p = 10^{-8}$ , and only 50% power to detect the lowest OR of 1.15 (*CD40*, MAF = 25%) at  $p = 0.05$ .

Results of our conditional analyses are shown in Table 3. Loci are shown in the order added by the algorithm (see Methods), i.e., rs261946 has the lowest  $p$ -value conditional on the SE, rs2074488 has the lowest  $p$ -value conditional on both SE and rs261946, and so on. Out of the 200 SNPs, three independent signals were evident in addition to the SE risk levels. One signal is located in the classical HLA class II region between genes *BTNL2* and *HLA-DRA*, and two signals are in the classical HLA class I region (see Figure 1) near *TRIM39* and *HLA-C*.

**Table 1: HLA-DRB1 risk levels. Definitions and case-control ratios for shared-epitope (SE) categories.**

DRB1 risk level	Definition	Case-control ratio
3 = Highest-risk SE	DRB1*0401, 0404, 0405, 0408, or 0409	72%-28%
2 = Lower-risk SE	Other SE alleles	52%-48%
1 = No SE	Not an SE allele	23%-77%

**Table 2: Significance of associations with published RA risk alleles**

SNP (GENE)	Reference	Proxy ( $r^2 = 1$ )	EIGENSTRAT adjusted $p$ -value
rs2476601 ( <i>PTPN22</i> )	[16]	N/A	$5.3 \times 10^{-6}$
rs7574865 ( <i>STAT4</i> )	[2]	N/A	0.034
rs3761847 ( <i>TRAF1-C5</i> )	[3]	N/A	$1.1 \times 10^{-5}$
rs1953126 ( <i>TRAF1-C5</i> )	[17]	N/A	$8.0 \times 10^{-6}$
rs10499194 ( <i>TNFAIP3</i> )	[18]	rs13192841	0.25
rs6920220 ( <i>TNFAIP3</i> )	[19]	rs6933404	0.17
rs4810485 ( <i>CD40</i> )	[20]	rs1569723	0.12

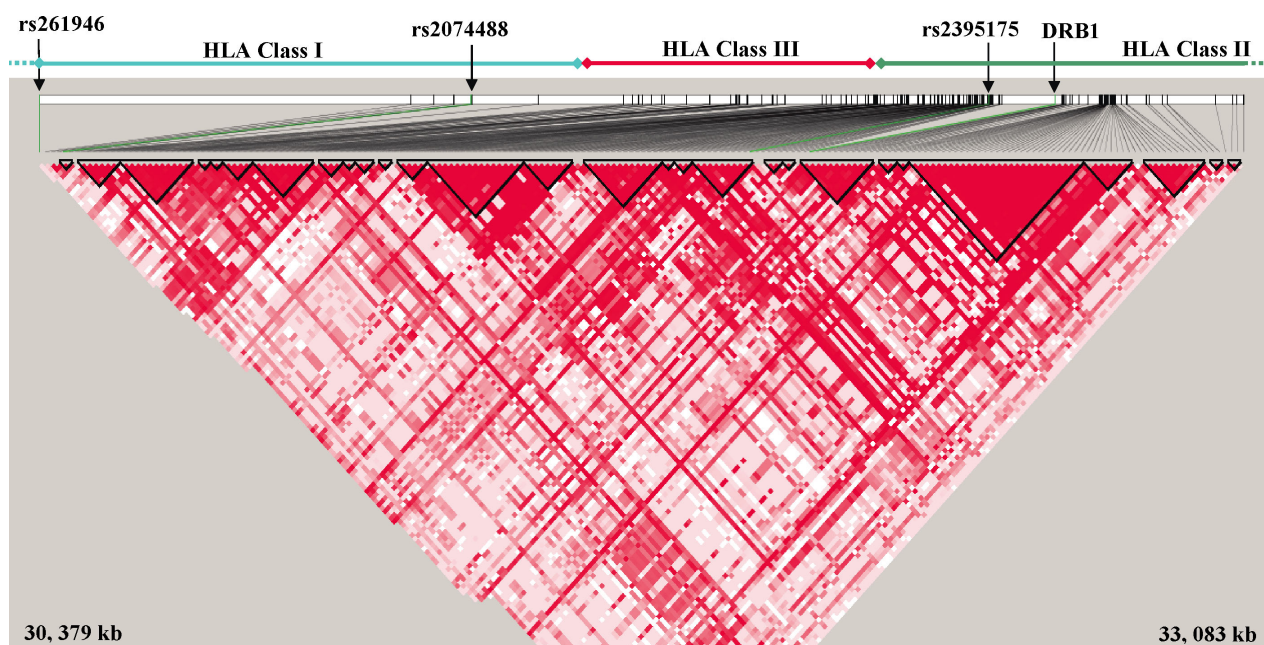
**Table 3: Independent SNPs in conditional analysis of RA susceptibility**

Locus (sequence #) <sup>a</sup>	Unadjusted single-marker <i>p</i> -value from haplview <sup>b</sup>	Single-marker EIGENSTRAT <i>p</i> -value (rank)	<i>p</i> -value conditional on loci above <sup>c</sup>	<i>p</i> -value conditional on other 3 loci <sup>c</sup>	Location (kb)	Closest gene(s)
Shared epitope (#129)	$1.9 \times 10^{-188}$	N/A	N/A	$8.6 \times 10^{-74}$	32,655-32,666	In DRB1
rs261946 (#1)	$7.2 \times 10^{-17}$	$1.9 \times 10^{-9}$ (185)	0.00019	0.0027	30,379	TRIM39
rs2074488 (#5)	$2.3 \times 10^{-24}$	$5.0 \times 10^{-12}$ (151)	$6.6 \times 10^{-5}$	0.0011	31,348	HLA-C
rs2395175 (#119)	$1.4 \times 10^{-117}$	$1.9 \times 10^{-80}$ (2)	0.00031	0.00031	32,513	Between BTNL2 and HLA-DRA

<sup>a</sup># indicates marker position in Figure 1. Loci are in order added by conditional algorithm (see Methods).

<sup>b</sup>For shared epitope, multi-allelic SNP test in Whap using No-Low-High risk.

<sup>c</sup>Whap log ratio test using all loci as alternate model versus null model without this locus.



**Figure 1**  
**Region with LD of SNPs studied in conditional analyses.** Haplotype diagram displays increasing red with higher  $D'$ . Three independently significant RA SNPs and *HLA-DRB1* locations are indicated.

Table 4 shows the case-control frequencies and ORs for the final haplotypes, with the most common haplotype (ACG-1) as the reference haplotype. The highest-risk haplotype of these SNPs and the SE level had OR = 27.2 (95% CI, 16.7-44.4) in comparison to OR = 7.3 (95% CI, 4.7-11.1) overall comparing the SE risk levels alone (data not shown).

**Conclusion**

In our analysis of the Genetic Analysis Workshop 16 dataset, there was insufficient power to detect known associations with RA susceptibility at a genome-wide significance level outside of the MHC; the most

significant association was  $p = 5.3 \times 10^{-6}$  for *PTPN22*. Clearly, the MHC is the most influential genetic region in RA susceptibility, but extensive LD makes isolating the precise loci difficult. We have used conditional analyses as a tool to investigate the presence of multiple RA risk factors in the MHC region in addition to the SE. Out of 200 candidate SNPs having unconditional  $p$ -values  $< 10^{-8}$ , we have identified an additional HLA class II marker and two HLA class I markers which have significant associations with RA susceptibility that are not fully explained by LD with *HLA-DRB1*. A better understanding of these genetic influences can be helpful in elucidating the complex genetic components of RA.

**Table 4: Frequencies and odds ratios for haplotypes of three SNPs and shared epitope proxy risk level**

Haplotype	Frequency in cases	Frequency in controls	OR (95% CI)
GAA-3	9.80%	1.90%	27.2 (16.7-44.4)
AAA-3	6.50%	1.50%	22.2 (12.6-39)
ACA-3	17.10%	4.40%	19.9 (13.4-29.7)
GCG-2	5.30%	2.40%	14.6 (8.6-24.8)
GCA-3	12.60%	4.20%	13.1 (9-19.1)
ACG-3	3.40%	1.80%	6.2 (3.7-10.4)
ACG-2	13.80%	9.50%	5.2 (3.8-7.1)
AAG-1	1.80%	2.70%	2.5 (1.4-4.5)
GAG-1	1.40%	2.40%	NS
GCG-1	6.00%	13.00%	NS
ACA-1	1.50%	2.50%	NS
ACG-1	20.80%	53.80%	(Reference group)

Previous studies of MHC effects on RA susceptibility beyond the SE have identified additional independent signals but have been largely inconsistent, due at least in part to the difficulty of narrowing down regions of association in the presence of extended LD [1,13]. Multiple studies have implicated the TNF-lymphotoxin locus in class III [1], which were not significant in our conditional analysis. Other studies also using NARAC cases have observed signals in class I [5,14], including *HLA-C*, our second SNP added in conditional analysis. Our first SNP is in the gene *TRIM39*, also in class I but not previously implicated. Our third SNP, in class II, is 150 kb upstream from *HLA-DRB1* between the *BTNL2* and *HLA-DRA* genes. *BTNL2* has been associated with RA, systemic lupus erythematosus, and type 1 diabetes [15]; this is attributed to its association with predisposing *HLA DQB1-DRB1* haplotypes, which may explain its presence in our data as well.

It is important to note that the NARAC population is primarily Caucasian. Other populations could have quite different distributions of these haplotypes as well as other haplotypes and allele frequencies. A similar analysis in other ethnic groups could be very informative.

### List of abbreviations used

LD: Linkage disequilibrium; MAF: Minor allele frequency; MHC: Major histocompatibility complex; NARAC: North American Rheumatoid Arthritis Consortium; OR: Odds ratio; PCA: Principal components analysis; RA: Rheumatoid arthritis; SE: Shared epitope; SNP: Single-nucleotide polymorphism.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

KET performed statistical analyses and drafted the manuscript. LAC recruited patients as part of the NARAC collaboration. LAC and KET designed the study, revised the manuscript, and read and approved the final manuscript.

### Acknowledgements

The Genetic Analysis Workshops are supported by NIH grant R01 GM031575 from the National Institute of General Medical Sciences. The authors are grateful for the support of Dr. Peter K. Gregersen and the NARAC investigative team.

This work was supported by grants from the National Institutes of Health: N01-AR-72232, R01 AI 065841, K24 AR02175, National Center for Research Resources M01 RR-00079 the Rosalind Russell Medical Research Center for Arthritis at UCSF, and a Kirkland Scholar Award to LAC.

This article has been published as part of *BMC Proceedings* Volume 3 Supplement 7, 2009: Genetic Analysis Workshop 16. The full contents of the supplement are available online at <http://www.biomedcentral.com/1753-6561/3?issue=S7>.

### References

- Newton JL, Harney SM, Wordsworth BP and Brown MA: **A review of the MHC genetics of rheumatoid arthritis.** *Genes Immun* 2004, **5**:151-157.
- Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PI, Le JM, Lee HS, Batiwalla F, Li W, Masters SL, Booty MG, Carulli JP, Padyukov L, Alfredsson L, Klareskog L, Chen WV, Amos CI, Criswell LA, Seldin MF, Kastner DL and Gregersen PK: **STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus.** *N Engl J Med* 2007, **357**:977-986.
- Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LR, Li W, Tan AK, Bonnard C, Ong RT, Thalamuthu A, Pettersson S, Liu C, Tian C, Chen WV, Carulli JP, Beckman EM, Altshuler D, Alfredsson L, Criswell LA, Amos CI, Seldin MF, Kastner DL, Klareskog L and Gregersen PK: **TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study.** *N Engl J Med* 2007, **357**:1199-1209.
- Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelicis M, Parkin M, Barry R, Winslow W, Healy C, Graham RR, Neale BM, Izmailova E, Roubenoff R, Parker AN, Glass R, Karlson EW, Maher N, Hafler DA, Lee DM, Seldin MF, Remmers EF, Lee AT, Padyukov L, Alfredsson L, Coby J, Weinblatt ME, Gabriel SB, Purcell S, Klareskog L, Gregersen PK, Shadick NA, Daly MJ and Altshuler D: **Two independent alleles at 6q23 associated with risk of rheumatoid arthritis.** *Nat Genet* 2007, **39**:1477-1482.
- Jawaheer D, Li W, Graham RR, Chen W, Damle A, Xiao X, Monteiro J, Khalili H, Lee A, Lundsten R, Begovich A, Bugawan T, Erlich H, Elder JT, Criswell LA, Seldin MF, Amos CI, Behrens TW and Gregersen PK: **Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis.** *Am J Hum Genet* 2002, **71**:585-594.
- Vignal C, Bansal AT, Balding DJ, Binks MH, Dickson MC, Montgomery DS and Wilson AG: **Genetic association of the major histocompatibility complex with rheumatoid arthritis implicates two non-DRB1 loci.** *Arthritis Rheum* 2009, **60**:53-62.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC: **PLINK: a toolset for whole-genome association and population-based linkage analysis.** *Am J Hum Genet* 2007, **81**:559-575.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA and Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet* 2006, **38**:904-909.
- Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, Silva G, Belmont JW, Klareskog L and Gregersen PK: **European population**

- substructure: clustering of northern and southern populations.** *PLoS Genet* 2006, **2**:1–13.
10. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A and Goldstein DB: **A whole-genome association study of major determinants for host control of HIV-1.** *Science* 2007, **317**:944–947.
  11. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PV, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapää-Dahlqvist S, Petri M, Manzi S, Seldin MF, Rönnblom L, Syvänen AC, Criswell LA, Gregersen PK and Behrens TW: **Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX.** *N Engl J Med* 2008, **358**:900–909.
  12. Purcell S, Daly M and Sham P: **WHAP: haplotype-based association analysis.** *Bioinformatics* 2007, **23**:255–256.
  13. Gregersen PK and Olsson LM: **Recent advances in the genetics of autoimmune disease.** *Annu Rev Immunol* 2009, **27**:363–391.
  14. Lee HS, Lee AT, Criswell LA, Seldin MF, Amos CI, Carulli JP, Navarrete C, Remmers EF, Kastner DL, Plenge RM, Li W and Gregersen PK: **Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus.** *Mol Med* 2008, **14**:293–300.
  15. Orozco G, Eerligh P, Sánchez E, Zhenakova S, Roep BO, González-Gay MA, López-Nevot MA, Callejas JL, Hidalgo C, Pascual-Salcedo D, Balsa A, González-Escribano MF, Koeleman BP and Martin J: **Analysis of a functional BTNL2 polymorphism in type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus.** *Hum Immunol* 2005, **66**:1235–1241.
  16. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoorke JM, Conn MT, Chang M, Chang SY, Saiki RK, Catanese JJ, Leong DU, Garcia VE, McAllister LB, Jeffery DA, Lee AT, Batliwalla F, Remmers E, Criswell LA, Seldin MF, Kastner DL, Amos CI, Sninsky JJ and Gregersen PK: **A missense single-nucleotide polymorphism in the protein tyrosine phosphatase PTPN22 is associated with rheumatoid arthritis.** *Am J Hum Genet* 2004, **75**:330–337.
  17. Kurreeman FA, Padyukov L, Marques RB, Schrodi SJ, Seddighzadeh M, Stoeken-Rijsbergen G, Helm-van Mil van der AH, Allaart CF, Verduyn W, Houwing-Duistermaat J, Alfredsson L, Begovich AB, Klareskog L, Huizinga TW and Toes RE: **A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis.** *PLoS Med* 2007, **4**:e278.
  18. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, Burt NP, Blumenstiel B, DeFelice M, Parkin M, Barry R, Winslow W, Healy C, Graham RR, Neale BM, Izmailova E, Roubenoff R, Parker AN, Glass R, Karlson EW, Maher N, Hafler DA, Lee DM, Seldin MF, Remmers EF, Lee AT, Padyukov L, Alfredsson L, Coby J, Weinblatt ME, Gabriel SB, Purcell S, Klareskog L, Gregersen PK, Shadick NA, Daly MJ and Altshuler D: **Two independent alleles at 6q23 associated with risk of rheumatoid arthritis.** *Nat Genet* 2007, **39**:1477–1482.
  19. Wellcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447**:661–678.
  20. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burt NP, Gianniny L, Korman BD, Padyukov L, Kurreeman FA, Chang M, Catanese JJ, Ding B, Wong S, Helm-van Mil van der AH, Neale BM, Coby J, Cui J, Tak PP, Wolbink GJ, Crusius JB, Horst-Bruinsma van der IE, Criswell LA, Amos CI, Seldin MF, Kastner DL, Ardlie KG, Alfredsson L, Costenbader KH, Altshuler D, Huizinga TW, Shadick NA, Weinblatt ME, de Vries N, Worthington J, Seielstad M, Toes RE, Karlson EW, Begovich AB, Klareskog L, Gregersen PK, Daly MJ and Plenge RM: **Common variants at CD40 and other loci confer risk of rheumatoid arthritis.** *Nat Genet* 2008, **40**:1216–1223.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

