DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Data for Genetic Analysis Workshop 16 Problem 1, Association Analysis of Rheumatoid Arthritis Data

The Harvard community has made this article openly available.
Please share how this access benefits you. Your story matters.

Citation	Amos, Christopher I., Wei Vivien Chen, Michael F. Seldin, Elaine F. Remmers, Kimberly E. Taylor, Lindsey A. Criswell, Annette T. Lee, Robert M. Plenge, Daniel L. Kastner, and Peter K. Gregersen. 2009. Data for Genetic Analysis Workshop 16 Problem 1, association analysis of rheumatoid arthritis data. BMC Proceedings 3(Suppl. 7): S2.
Published Version	doi:10.1186/1753-6561-3-s7-s2
Accessed	February 19, 2015 7:11:00 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:10139320
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

Proceedings

Data for Genetic Analysis Workshop 16 Problem 1, association analysis of rheumatoid arthritis data

Christopher I Amos*[1], Wei Vivien Chen[1], Michael F Seldin[2], Elaine F Remmers[3], Kimberly E Taylor[4], Lindsey A Criswell[4], Annette T Lee[5], Robert M Plenge[6], Daniel L Kastner[3] and Peter K Gregersen[5]

Addresses: [1]Departments of Epidemiology and Biomathematics, University of Texas, MD Anderson Cancer Center, 1155 Pressler Street, Houston, Texas 77030, USA, [2]Rowe Program of Human Genetics, Department of Medicine, University of California, One Shields Avenue, 4303 Tupper Hall, Davis, California 95616, USA, [3]Genetics and Genomics Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, 9000 Rockville Pike, Building 10, Room 6N226, Bethesda, Maryland 20892, USA, [4]The Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, Division of Rheumatology, University of California at San Francisco, 374 Parnassus Avenue, Box 0500, San Francisco, California 94143-0500, USA, [5]Center for Genomics and Human Genetics, North Shore-Feinstein Medical Research Institute, 350 Community Drive, Manhasset, New York 11030, USA and [6]Program in Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology and Harvard, 320 Charles Street, Cambridge, Massachusetts 02141, USA

E-mail: Christopher I Amos* - camos@mdanderson.org; Wei Vivien Chen - weichen@mdanderson.org; Michael F Seldin - mfseldin@ucdavis.edu; Elaine F Remmers - remmerse@mail.nih.gov; Kimberly E Taylor - ktaylor@medicine.ucsf.edu; Lindsey A Criswell - lindsey.criswell@ucsf.edu; Annette T Lee - Anlee@nshs.edu; Robert M Plenge - rplenge@partners.org; Daniel L Kastner - kastnerd@exchange.nih.gov; Peter K Gregersen - peterg@nshs.edu
*Corresponding author

This article is available from: http://www.biomedcentral.com/1753-6561/3/S7/S2

Abstract

For Genetic Analysis Workshop 16 Problem 1, we provided data for genome-wide association analysis of rheumatoid arthritis. Single-nucleotide polymorphism (SNP) genotype data were provided for 868 cases and 1194 controls that had been assayed using an Illumina 550 k platform. In addition, phenotypic data were provided from genotyping DRB1 alleles, which were classified according to the rheumatoid arthritis shared epitope, levels of anti-cyclic citrullinated peptide, and levels of rheumatoid factor IgM. Several questions could be addressed using the data, including analysis of genetic associations using single SNPs or haplotypes, as well as gene-gene and genetic analysis of SNPs for qualitative and quantitative factors.

Background

Rheumatoid arthritis is a complex disease with a moderately strong genetic component. The recurrence risk ratio for siblings is typically estimated at around 6 in Caucasians, but it has a broad range of values, primarily because the prevalence in the population is not well characterized [1]. The prevalence also varies among populations, ranging from around 0.8% in Caucasians to

10% in some Native American groups. Females are generally at higher risk than males, with about a 3 to 1 predominance of females to males. The mean age of disease onset is in the fifth decade with considerable variability in age at presentation, including occasional presentation in the teenage years.

The HLA region on 6p21 has been implicated by numerous studies, and there is consistent evidence that DR alleles contribute to disease risk. The 'shared epitope' hypothesis was proposed by Gregersen et al. [2] to explain the organization of risk for rheumatoid arthritis from DR alleles. According to this hypothesis, individuals who share a QK/RRAA motif in positions 70 to 74 of the DR molecule show an increased risk for disease. The alleles that confer increased risk for rheumatoid arthritis include *DRB1*0101*, *0102*, *0104*, *0105*, **0401**, **0404**, **0405**, **0408**, **0409**, *1001*, *1402*, and *1406*, with highest risk alleles in bold [3]. This model was not quite sufficient to explain risk according to DR types, and newer models utilizing data from positions 70 to 74 have been developed [4,5]. DR effects on risk for rheumatoid arthritis also show a complex effect on risk for rheumatoid arthritis, but presence of two risk alleles generally increases risk substantially more than the risk associated with heterozygosity for risk and nonrisk alleles. Aside from the main effects of DR, there is also evidence for interactions with other HLA loci or haplotypic effects including the class 1 region and the central MHC [6]. Certain DR alleles, notably DR3 [7,8], can occur on a background of extended linkage disequilibrium, for which the extended haplotype confers increased risk, even though DR3 alleles alone do not increase risk.

Two quantitative phenotypes that are used for identifying rheumatoid arthritis affected individuals include anti-cyclic citrullinated peptide (anti-CCP) and rheumatoid factor IgM autoantibodies. The heritability of these measures is hard to obtain from the selected sib pairs we are studying. After proband correction, the heritability estimates are 11% and 30%, while before correction the heritabilities are 15% and 67%. Specific autoantibodies are noted to co-occur with rheumatoid arthritis. Rheumatoid factor IgM has been correlated with erosive arthritic disease. However, anti-CCP is more specific for the disease and is a better predictor of erosive outcome [9]. Elevations of anti-CCP have been noted to predict increased risk for development of rheumatoid arthritis [10]. The shared-epitope alleles are strongly associated with the presence of anti-CCP antibodies, and there is evidence that this effect is modulated by HLA-DR3 [8].

Alleles at the *PTPN22* locus have been shown to confer an increased risk for rheumatoid arthritis [11]. At least two alleles of *PTPN22* have been implicated as causing increased risk for rheumatoid arthritis; the R620W allele in rs2476601 (hCV16021387) confers 1.7- to 1.9-fold increased risk to heterozygotes and higher risks to homozygous carriers. These findings have further been confirmed by analysis of transmission of *PTPN22* alleles to affected offspring in families [12]. Increased risk has also been noted for either hCV8689108 or hCV25762283 [13], with some indeterminacy because of linkage disequilibrium among these markers (and others in the region).

The *CTLA4* locus on chromosome 2q33 has been associated with mildly increased risk for rheumatoid arthritis [14]. In addition, alleles at loci in the *TRAF1/C5* region are associated with rheumatoid arthritis risk [15]. A targeted association study showed that alleles of *STAT4* [16] are associated with rheumatoid arthritis risk, but these associations are too weak to reach genome-wide levels of association in the data set that we have here provided. Similarly, a locus on chromosome 6q (*TNFAIP3*) that is associated with rheumatoid arthritis risk has relatively weaker effects [15]. Additional loci that have been implicated in Caucasian rheumatoid arthritis populations include *CD40* (20q13), *PRKCQ* (10p15), and *CCL21* (9p13), among others [17,18].

Aside from identified genetic factors and sex, few environmental cofactors have been identified as affecting risk for rheumatoid arthritis. However, current smoking confers about a two-fold increased risk [7]. Klareskog et al. [19] showed that the risk from smoking for rheumatoid arthritis is particularly high among individuals who have a shared-epitope allele and who also have elevated levels of anti-CCP antibodies. The biological basis for this rather complex interaction appears to reflect increased citrullination of peptides among smokers, and presentation of citrullinated peptides by shared-epitope alleles.

The data set submitted for the Genetic Analysis Workshop 16 (GAW16) was designed with a primary goal of allowing the identification of genetic factors that predispose to rheumatoid arthritis using association methods. Given some previously identified evidence for effects of smoking on rheumatoid arthritis risk and difference in risk according to sex, there is an interest in identifying gene-environment and gene-gene combinations that yield particularly high risks to individuals for rheumatoid arthritis.

Methods
The cases that we made available for analysis by participants in GAW16 comprised independent

individuals who had met the American College of Rheumatology criteria for rheumatoid arthritis. These cases comprise a single member of 445 sibpairs that were studied as a part of the North American Rheumatoid Arthritic Consortium because they had at least one additional sibling with rheumatoid arthritis, and an additional 423 cases who were not selected for family history. The cases were recruited from across the United States. Cases are predominantly of Northern European origin. The controls, derived from the New York Cancer Project, were enrolled in the New York metropolitan area [20]. These controls are somewhat enriched for individuals of Southern European or Ashkenazi Jewish ancestry compared with cases. Structure across European populations has been described [21,22], and some autoimmune predisposing alleles, such as the *PTPN22* R620W and *HLA DR4* alleles show strong clines across European populations. In addition, alleles at other loci such as the Lactase Persistence gene (*LCT*) show strong clines across European populations. Evidence in association studies for an effect of the *LCT* locus on case/control status likely reflects false-positive association due to stratification. Studies within Europe have confirmed the associations of *PTPN22* and *HLA* but have not confirmed effects of *LCT* on risk for rheumatoid arthritis.

The GAW16 rheumatoid arthritis data is part of ongoing studies to identify genetic associations of rheumatoid arthritis [14]. The data that were provided to GAW16 included results from genotyping 868 cases and 1194 controls after the application of quality control procedures that included removing individuals who had a low overall call rate (<95%) of single-nucleotide polymorphisms (SNPs), removing first degree relatives, and removing duplicated and contaminated samples. The data that were provided as a part of Genetic Analysis Workshop 16 Problem 1 were included in a previous publication [15], which identified the *TRAF1/C5* locus as contributing to susceptibility to rheumatoid arthritis. This earlier publication included additional data that were not provided to the Genetic Analysis Workshop 16 Problem 1 from a study of early-onset rheumatoid arthritis conducted in Sweden. Aside from the *TRAF1/C5* locus, there were significant effects from the HLA region and *PTPN22* that can be readily discerned from the data.

Data that were provided to Genetic Analysis Workshop 16 participants included affection status with rheumatoid arthritis, sex, *DRB1* alleles detected by serology and further defined using DNA probes for *DRB1*04* and *DRB1*01* alleles, number of shared epitopes carried, the anti-CCP titer, rheumatoid factor IgM level, and 545,080 genotypes derived from Illumina genotyping arrays. All rheumatoid arthritis cases and 589 controls were genotyped on the HumanHap500 v1, 358 controls

were done on the Human Hap500 v3.0, and 247 controls were done on HumanHap300 and HumanHap240 arrays.

Discussion
Rheumatoid arthritis results from a complex interaction of genetic and environmental factors. Data that were provided for GAW16 were derived from a large number of cases and controls who had been genotyped using dense SNP arrays. These data were sufficient to identify many genetic loci influencing rheumatoid arthritis risk. In addition, we provided data for two autoantibodies that are often elevated among individuals who have rheumatoid arthritis. Aside from identifying genetic factors influencing rheumatoid arthritis, the data that were provided can be used to investigate population structure in European populations, methods for inferring SNPs, and modeling approaches when multiple genetic factors influence disease risk.

List of abbreviations used
anti-CCP: Anti-cyclic citrullinated peptide; GAW: Genetic Analysis Workshop; SNP: Single-nucleotide polymorphism.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
CIA developed data to be transmitted and wrote the first and final drafts of the submitted manuscript. WVC assembled data to be transmitted and performed summary analyses. MFS, ER, LAC, ATL, DLK, and PKG have participated in development of clinical and genetic data. KET, RMP, MFS, and PKG provided input in organizing analyses and in interpretation of results. CIA, WVC, MFS, ER, LAC, RMP, and PKG provided assistance in manuscript preparation. All authors read and approved the final manuscript.

References

1. Seldin MF, Amos CI, Ward R and Gregersen PK: **The genetics revolution and the assault on rheumatoid arthritis.** *Arthritis Rheum* 1999, **42:**1071–1079.
2. Gregersen PK, Silver J and Winchester RJ: **The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis.** *Arthritis Rheum* 1987, **30:**1205–1213.
3. Newton JL, Harney SM, Wordsworth BP and Brown MA: **A review of the MHC genetics of rheumatoid arthritis.** *Genes Immun* 2004, **5:**151–157.
4. du Montcel ST, Michou L, Petit-Teixeira E, Osorio J, Lemaire I, Lasbleiz S, Pierlot C, Quillet P, Bardin T, Prum B, Cornelis F and Clerget-Darpoux F: **New classification of HLA-DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility.** *Arthritis Rheum* 2005, **52:**1063–1068.
5. Morgan AW, Haroon-Rashid L, Martin SG, Gooi HC, Worthington J, Thomson W, Barrett JH and Emery P: **The shared epitope hypothesis in rheumatoid arthritis: evaluation of alternative classification criteria in a large UK Caucasian cohort.** *Arthritis Rheum* 2008, **58:**1275–1283.
6. Lee HS, Korman BD, Le JM, Kastner DL, Remmers EF, Gregersen PK and Bae SC: **Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations.** *Arthritis Rheum* 2009, **60:**364–371.
7. 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 and 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.
8. Irigoyen P, Lee AT, Wener MH, Li W, Kern M, Batliwalla F, Lum RF, Massarotti E, Weisman M, Bombardier C, Remmers EF, Kastner DL, Seldin MF, Criswell LA and Gregersen PK: **Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles.** *Arthritis Rheum* 2005, **52:**3813–3818.
9. Huizinga TW, Amos CI, Helm-van Mil van der AH, Chen W, van Gaalen FA, Jawaheer D, Schreuder GM, Wener M, Breedveld FC, Ahmad N, Lum RF, de Vries RR, Gregersen PK, Toes RE and Criswell LA: **Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins.** *Arthritis Rheum* 2005, **52:**3433–3438.
10. Kroot EJ, de Jong BA, van Leeuwen MA, Swinkels H, Hoogen van den FH, van't Hof M, Putte van de LB, van Rijswijk MH, van Venrooij WJ and van Riel PL: **The prognostic value of anti-cyclic citrullinated peptide antibody in patients with recent-onset rheumatoid arthritis.** *Arthritis Rheum* 2000, **43:**1831–1835.
11. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoerke 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 a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis.** *Am J Hum Genet* 2004, **75:**330–337.
12. Michou L, Lasbleiz S, Rat AC, Migliorini P, Balsa A, Westhovens R, Barrera P, Alves H, Pierlot C, Glikmans E, Garnier S, Dausset J, Vaz C, Fernandes M, Petit-Teixeira E, Lemaire I, Pascual-Salcedo D, Bombardieri S, Dequeker J, Radstake TR, Van Riel P, Putte van de L, Lopes-Vaz A, Prum B, Bardin T, Dieudé P, Cornélis F and European Consortium on Rheumatoid Arthritis Families: **Linkage proof for PTPN22, a rheumatoid arthritis susceptibility gene and a human autoimmunity gene.** *Proc Natl Acad Sci USA* 2007, **104:**1649–1654.
13. Carlton VE, Hu X, Chokkalingam AP, Schrodi SJ, Brandon R, Alexander HC, Chang M, Catanese JJ, Leong DU, Ardlie KG, Kastner DL, Seldin MF, Criswell LA, Gregersen PK, Beasley E, Thomson G, Amos CI and Begovich AB: **PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid arthritis.** *Am J Hum Genet* 2005, **77:**567–581.
14. Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, Wolfe F, Kastner DL, Alfredsson L, Altshuler D, Gregersen PK, Klareskog L and Rioux JD: **Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4.** *Am J Hum Genet* 2005, **77:**1044–1060.
15. 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.
16. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PI, Le JM, Lee HS, Batliwalla 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.
17. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burtt 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, Coblyn 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.
18. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, Plant D, Gibbons LJ, Wellcome Trust Case Control Consortium; YEAR Consortium; BIRAC Consortium, Wilson AG, Bax DE, Morgan AW, Emery P, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P and Worthington J: **Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13.** *Nat Genet* 2008, **40:**1156–1159.
19. Klareskog L, Stolt P, Lundberg K, Källberg H, Bengtsson C, Grunewald J, Rönnelid J, Harris HE, Ulfgren AK, Rantapää-Dahlqvist S, Eklund A, Padyukov L and Alfredsson L: **A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination.** *Arthritis Rheum* 2006, **54:**38–46.
20. Mitchell MK, Gregersen PK, Johnson S, Parsons R and Vlahov D: **The New York Cancer Project: rationale, organization, design, and baseline characteristics.** *J Urban Health* 2004, **81:**301–310.
21. 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:**e143.
22. Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C, Klareskog L, Pulver AE, Qi L, Gregersen PK and Seldin MF: **Analysis and application of European genetic substructure using 300 K SNP information.** *PLoS Genet* 2008, **4:**e4.