

Several Regions in the Major Histocompatibility Complex Confer Risk for Anti-CCP-Antibody Positive Rheumatoid Arthritis, Independent of the DRB1 Locus

Hye-Soon Lee,^{1,2} Annette T Lee,¹ Lindsey A Criswell,³ Michael F Seldin,⁴ Christopher I Amos,⁵ John P Carulli,⁶ Cristina Navarrete,⁷ Elaine F Remmers,⁸ Daniel L Kastner,⁸ Robert M Plenge,⁹ Wentian Li,¹ and Peter K Gregersen¹

¹The Feinstein Institute for Medical Research, North Shore LIJ Health System, Manhasset, New York, United States of America; ²Hanyang University College of Medicine, Seoul, South Korea; ³The University of California San Francisco, San Francisco, California, United States of America; ⁴the University of California Davis, Davis, California, United States of America; ⁵The University of Texas, M.D. Anderson Cancer Center, Houston, Texas, United States of America; ⁶Biogen Idec Inc., Cambridge, Massachusetts, United States of America; ⁷National Blood Service, London, United Kingdom; ⁸The National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, Maryland, United States of America; ⁹The Broad Institute of Harvard and the Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America; ¹⁰Division of Rheumatology, Immunology and Allergy, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

Recent evidence suggests that additional risk loci for RA are present in the major histocompatibility complex (MHC), independent of the class II HLA-DRB1 locus. We have now tested a total of 1,769 SNPs across 7.5Mb of the MHC located from 6p22.2 (26.03 Mb) to 6p21.32 (33.59 Mb) derived from the Illumina 550K Beadchip (Illumina, San Diego, CA, USA). For an initial analysis in the whole dataset (869 RA CCP + cases, 1,193 controls), the strongest association signal was observed in markers near the HLA-DRB1 locus, with additional evidence for association extending out into the Class I HLA region. To avoid confounding that may arise due to linkage disequilibrium with DRB1 alleles, we analyzed a subset of the data by matching cases and controls by DRB1 genotype (both alleles matched 1:1), yielding a set of 372 cases with 372 controls. This analysis revealed the presence of at least two regions of association with RA in the Class I region, independent of DRB1 genotype. SNP alleles found on the conserved A1-B8-DR3 (8.1) haplotype show the strongest evidence of positive association ($P \sim 0.00005$) clustered in the region around the HLA-C locus. In addition, we identified risk alleles that are not present on the 8.1 haplotype, with maximal association signals ($P \sim 0.001-0.0027$) located near the ZNF311 locus. This latter association is enriched in DRB1*0404 individuals. Finally, several additional association signals were found in the extreme centromeric portion of the MHC, in regions containing the DOB1, TAP2, DPB1, and COL11A2 genes. These data emphasize that further analysis of the MHC is likely to reveal genetic risk factors for rheumatoid arthritis that are independent of the DRB1 shared epitope alleles.

Online address: <http://www.molmed.org>
doi: 10.2119/2007-00123.Lee

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic autoimmune inflammatory disease characterized by joint inflammation and progressive joint destruction (1). Recently, several new genes with modest levels of risk for RA have been identified and replicated in various populations, includ-

ing PTPN22 (2), PADI4 (3), TRAF1-C5 (4, 5), STAT4 (5), and 6q23/TNFAIP3 (6). Nevertheless, numerous association studies and multiple genomewide linkage studies have shown that MHC region has the largest and most consistent genetic contribution in rheumatoid arthritis (7-9).

Because the association of HLA with RA was first demonstrated in 1976 (10), the vast majority of case-control association studies have focused on HLA-DRB1 locus encoding a group of risk alleles collectively called the "shared epitope" (SE) alleles (11,12). These alleles share a common sequence element containing Q/K-R-R-A-A at positions 70-74 of the DRB1 chain, with some minor variation from this canonical sequence in some risk alleles. Despite the appealing simplicity of the shared epitope as an explanation for disease association, it is quite apparent that there is a complex hierarchy of risk for the various DRB1 alleles

Address correspondence and reprint requests to Peter K Gregersen, Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore LIJ Health System, 350 Community Drive, Manhasset, NY 11030. Phone: 516-562-1542; Fax: 516-562-1153; E-mail: peterg@nshs.edu.

Submitted November 26, 2007. Accepted for publication February 18, 2008; Epub (www.molmed.org) ahead of print February 20, 2008.

that contain the shared epitope (13). In addition, certain genotypic combinations, such as DRB1*0401/0404, carry exceedingly high risk that cannot be explained simply by the number of shared epitope alleles that are present (12). This suggests that there may be haplotypic effects that modify the risk of particular shared epitope alleles.

In addition, although the DRB1 locus is clearly of predominant importance, several reports over the years have suggested the presence of additional risk loci within the MHC (14-17). The arguments for these additional loci are often confounded by the complex patterns of linkage disequilibrium that are observed in this genetic region. By carefully matching cases and controls by DRB1 genotype, we now provide additional evidence for several new risk loci for RA located in the Class I region of the MHC, as well in the region centromeric to the DRB1 locus.

MATERIALS AND METHODS

Study Populations

RA cases and controls in the current analysis are taken largely from populations utilized for our previous whole genome association study using the Illumina 550K Beadchip (Illumina) (4). Briefly, RA cases were selected from four North American RA patient collections. The North American Rheumatoid Arthritis Consortium (NARAC) samples are from multiplex families (primarily affected sibling pairs); at least one sibling was required to have documented erosions on hand radiographs, with at least one sibling having disease onset between the ages of 18 and 60 years of age (18). The other collections include samples from the Wichita Rheumatic Disease Data Bank (WRDDB) (19), mean disease duration ten years; the National Inception Cohort of Rheumatoid Arthritis Patients (NICRAP) (20), enrolled within six months of clinical diagnosis; and Study of New Onset Rheumatoid Arthritis (SONORA) (21), enrolled within 3-12 months of clinical diagnosis. All cases were anticyclic citrullinated peptide anti-

body positive (CCP +) with reported European-American ancestry. The controls were taken from 1,732 individuals who are part of the New York Cancer Project (NYCP) (22) and on whom HLA-DRB1 data was available. All subjects reported European American ancestry. For the matched case-control studies, an additional set of 46 controls from the UK carrying the DRB1*0401/0404 genotype were included in the analysis. Informed consent was obtained for all samples using protocols approved by the local institutional review boards.

MHC Genotyping

Genotype data was obtained from the Illumina HumanHap550 genotyping array (Illumina) and included 2,094 SNPs in 7.56 Mb region from 6p22.2 (26.03 Mb) to 6p21.32 (33.59Mb) encompassing the entire MHC. Genotyping was performed at the Feinstein Institute for Medical Research according to the Illumina Infinium two assay manual (Illumina), as previously described (4). After stringent filtering with removal of SNPs with > 2% of missing genotype a total of 1,769 SNPs were available for analysis.

All participants were HLA-DRB1 typed using the SSOP low-resolution method (23); individuals with DRB1*04 and DRB1*01 were subsequently tested using medium-resolution panel to allow for four digit DRB1 subtyping. A proportion of the subjects had four digit typing for all DRB1 alleles (*see Supplementary Materials*). The 46 UK controls (DRB1*0401/0404 heterozygotes) were typed by the PCRrSSO technique using RELI™ kits (DYNAL).

Among an initial set of case samples for whole genome association study (total $n = 908$, including 464 NARAC, 168 WRDDB, 162 NICRAP, and 114 SONORA), we first carried out genome-wide quality control filtering as follows: individuals with > 5% missing genotypes, SNPs with > 5% missing data, control HWE P -values < 0.00001, and MAF < 0.01 were excluded. We removed samples with excess sharing across all pairs of individuals, as this pattern is consistent with DNA

sample contamination. We also have removed genetic outliers as determined by either 235 major ancestral informative SNPs or 1,411 European ancestral informative SNPs (24), using the first or second principal components (*see Supplementary Materials*). After filtering, 869 case samples and 1,193 controls remained for analysis. Of these, 700 cases and 1,172 controls had DRB1 oligotyping data available. Of these 700 cases, 74 are overlapping with NARAC cases that have been previously utilized for fine mapping studies in the MHC (16).

Statistical Analysis

We first conducted a case-control association study for single marker analysis using HelixTree 5.0.2 software (Golden Helix Inc, Bozhan, MT, USA) and R program, using all samples ($n = 869$ cases, 1,193 controls) and MHC SNPs ($n = 1,769$) that passed quality control filters. We calculated Pearson's chi-square test based on allele counts.

We further characterized association signals according to whether SNP alleles are found on the conserved ancestral A1-B8-DR3 (8.1) haplotype (25). Specifically, we distinguished association signals that were due to a higher 8.1 related allele frequency in the case group versus signals characterized by a higher 8.1 allele frequency in the control group. To determine whether particular alleles are present on the A1-B8-DR3 (8.1) ancestral haplotype, we obtained the sequence of the 8.1 haplotype from complete genomic sequences of 8.1 homozygotes (COX cell line) (<http://www.sanger.ac.uk/HGP/Chr6/MHC>). As an independent check, we used 20 samples (one case and 19 controls) from DRB1*0301 homozygotes among the present study population to identify the alleles of 8.1 haplotype. By definition, these samples carry two copies of the 8.1 haplotype at the DRB1 locus, and due to the long-range nature of the 8.1 haplotype, they also carry two copies of the 8.1 allele at nearby SNPs until the haplotype is broken by recombination. Because recombination at any given position only occurred in a fraction of these sam-

ples, the majority of the 20 samples had extended regions of SNP homozygosity reflecting the 8.1 haplotype. This can be used to infer the 8.1 haplotype associated alleles for each SNP.

To remove the effect of DRB1, the predominant risk locus for RA, we constructed a matched dataset with equal number of case and control samples with the exactly same DRB1 genotype. For matching of cases and controls at the DRB1 locus, we had available a mix of two digit (low resolution) for the majority of DRB1 alleles and four digit (medium resolution) oligotyping for DRB1*04 alleles and a large fraction of DRB1*01 related alleles. The vast majority of DRB1*01 alleles with four digit genotyping were DRB1*0101. Thus, matching on DRB1*04 and DRB1*01 subtypes was quite precise, while matching on other DRB1 genotypes was often based on membership in a common DRB1 allelic group. However, four digit typing also was available for some of these other allelic groups. The complete list of the DRB1 genotypes found in the matched pairs is given in Table 1 of the Supplementary Material. Because of the large number of DRB1*0401/DRB1*0404 heterozygotes in cases, we included additional controls with this genotype from the National Blood Service, UK.

In addition to matching of cases and controls by DRB1 genotype, we also carried out principal components analysis (EIGENSTRAT) (26) to assure matching of cases and controls for ancestry, thereby minimizing population stratification unrelated to disease status. Because the population stratification due to major ancestry differences was achieved by the removal of outliers during the quality control stage, the matching of case and control's ancestry was mainly on the European subpopulation level using 1,441 non-MHC makers selected to capture European population substructure (24). Therefore, case and control matched pairs were chosen not only by having the same DRB1 genotype, but also by virtue of proximity to each other in the PC1-PC2 plane (distance < 0.05) using

the 1,441 above mentioned SNPs (see figure 1 of Supplementary Materials). If no such pair could be found, they were excluded. This led to the selection 744 matched samples (372 pairs).

For exploratory subset analyses, we characterized DRB1 alleles into three groups according to SE status and DRB1 family group: The SE(a) group alleles contained 0401, 0404, 0405, and 0408. The SE(b) group alleles contained 0101, 0102, 0901, and 1001; the SE(-) group contained all other DRB1 alleles. For some analyses we selected particular genotypic subgroups including 0401/not 0404, or 0404/not 0401.

All supplementary materials are available online at molmed.org

RESULTS

Case-Control Association Study of SNPs Across the MHC Region in the Entire Dataset

To assess the patterns of association across the MHC, we analyzed our panel of MHC SNP markers extending from 6p22.2 (26.03 Mb) to 6p21.32 (33.59 Mb). These data were derived from a previous whole genome study using the Illumina 550K Beadchip (Illumina) (4). After quality filtering (see Methods), 1,769 SNPs were available for analysis on 869 CCP + case samples and 1,193 controls.

As expected, the strongest association signal was observed in markers near to the HLA-DRB1 locus (see Figure 1, which shows the single-SNP chi-square test statistic from the 2-by-2 allele count table). Nevertheless, there are several regions across the MHC, especially within the central MHC and Class I regions, that also showed significant association with RA (chi-square values between 50 and 100). Additional association signals of this magnitude are also observed in the region just centromeric to the DRB1 locus. It is, of course, unclear from this analysis whether these additional signals simply reflect linkage disequilibrium with the known DRB1 risk alleles, or whether these are independent effects.

Analysis of Case Control Pairs Matched 1:1 by DRB1 Genotype

To control for the effects of LD due to the strong association signals at the DRB1 locus, we chose to select an equal number of cases and controls matched specifically by DRB1 genotype, thereby eliminating differences between cases and controls at this locus. This approach is more precise than matching cases and controls by the simple presence of risk alleles and non risk alleles, because it is well established that there is considerable heterogeneity in the strength and evidence of association at the DRB1 locus in these two categories (12).

We were able to select 372 cases and 372 controls that were matched at the DRB1 locus for both alleles. As described in the Methods section, we also took care to match each pair of cases and controls for ancestry, using a panel of ancestry informative markers (24) and selecting pairs to minimize the distance between pairs using a principal components approach. Obviously, this dataset is considerably smaller than the sum of available samples. For example, the various SE(-)/SE(-) combinations of DRB1 alleles are in vast excess in controls, compared with the small number of cases available for matching on these genotypes. Conversely, some risk genotypes (such as 0401/0401 or 0401/0404) are rather uncommon in control subjects, and thus not all cases with this risk genotype could be utilized for matching. However, at this stage of the analysis, we preferred to remove any potential cause of a false positive signal rather than gain statistical power. To partially address the lack of controls with these high risk genotypes, we specifically obtained a group of 46 DRB1*0401/0404 controls to increase our sample size.

Our approach to the analysis was also guided by our previous studies that implicated the A1-B8-DR3 (8.1) haplotype in contributing to MHC risk that is independent of the DRB1 locus (16). Therefore, we categorized all 1769 SNPs as to whether the allele that is found on the 8.1 haplotype is positively or negatively

Table 1. List of markers with $\chi^2 > 6.635$ in 372 RA cases and 372 controls matched by DRB1 genotype. Odds ratios are shown assuming the risk allele is on the ancestral 8.1 haplotype (see Figure 2).

Marker	Position(bp)	Gene	8.1AF_Control ^a	8.1AF_Case ^b	χ^2	P-value	OR	OR 95%CI
rs1341758	26308498	HIST1H2BF	0.962	0.93	6.99	0.0082	0.52	(0.32,0.83)
rs6913828	26354644	HIST1H4G	0.961	0.925	8.43	0.0037	0.5	(0.31,0.79)
rs3763338	29002291	RFP	0.935	0.891	8.6	0.0034	0.57	(0.39,0.82)
rs2071788	29020287	RFP	0.937	0.898	6.95	0.0084	0.59	(0.41,0.87)
rs6923005	29084051	ZNF311	0.77	0.7	8.98	0.0027	0.7	(0.55,0.88)
rs6930903	29089224	ZNF311	0.77	0.7	8.98	0.0027	0.7	(0.55,0.88)
rs6456889	29118804	OR2W1	0.78	0.719	6.93	0.0085	0.72	(0.57,0.92)
rs9295794	29212915	OR2J3	0.94	0.902	6.66	0.0098	0.59	(0.4,0.87)
rs9357078	29405646	OR5U1	0.923	0.88	7.21	0.0073	0.61	(0.43,0.87)
rs9380120	29443517	OR12D3	0.921	0.879	6.71	0.0096	0.63	(0.44,0.88)
rs2535315	31160107	C6orf15	0.58	0.648	7.11	0.0077	1.34	(1.09,1.66)
rs3778638	31200104	PSORS1C1	0.768	0.84	11.56	0.00067	1.58	(1.22,2.05)
rs3130453	31232829	CCHCR1	0.425	0.5	8.18	0.0042	1.35	(1.1,1.66)
rs1419881	31238573	TCF19	0.421	0.489	6.78	0.0092	1.32	(1.07,1.62)
rs887464	31253900	POU5F1	0.409	0.484	8.21	0.0042	1.36	(1.11,1.67)
rs3130531	31314596	HLA-C	0.575	0.653	9.15	0.0025	1.39	(1.13,1.72)
rs2844623	31340523	HLA-C	0.753	0.826	11.15	0.00084	1.56	(1.21,2)
rs2524077	31351583	HLA-C	0.783	0.863	15.55	8.10E-05	1.75	(1.33,2.29)
rs3130696	31351864	HLA-C	0.192	0.251	7.29	0.007	1.43	(1.11,1.84)
rs2524073	31352215	HLA-C	0.793	0.87	15.03	0.00011	1.75	(1.32,2.31)
rs2524044	31364733	HLA-C	0.762	0.824	8.29	0.004	1.47	(1.14,1.89)
rs3873385	31377288	HLA-C	0.77	0.832	8.6	0.0034	1.48	(1.15,1.92)
rs9264868	31379581	HLA-C	0.79	0.871	16.44	5.00E-05	1.79	(1.36,2.37)
rs9264869	31379610	HLA-C	0.793	0.87	15.03	0.00011	1.75	(1.32,2.31)
rs9264904	31380533	HLA-C	0.665	0.749	12.26	0.00046	1.51	(1.2,1.89)
rs1634747	31389856	HLA-C	0.402	0.485	10.04	0.0015	1.41	(1.14,1.73)
rs2596560	31463298	MICA	0.208	0.276	9.03	0.0027	1.46	(1.14,1.85)
rs3099844	31556956	HCP5	0.083	0.125	6.69	0.0097	1.6	(1.13,2.25)
rs2855812	31580700	MICB	0.226	0.289	7.44	0.0064	1.4	(1.1,1.77)
rs2857595	31676449	NCR3	0.124	0.183	9.44	0.0021	1.59	(1.19,2.12)
rs9267658	31953965	SLC44A4	0.876	0.919	6.86	0.0088	1.61	(1.14,2.27)
rs3134954	32179872	TNXB	0.872	0.918	7.8	0.0052	1.65	(1.17,2.33)
rs1035798	32259201	AGER	0.808	0.746	7.84	0.0051	0.7	(0.55,0.89)
rs394657	32295002	NOTCH4	0.342	0.417	8.47	0.0036	1.39	(1.12,1.72)
rs2857210	32849721	HLA-DQB2	0.682	0.767	13.28	0.00027	1.55	(1.23,1.96)
rs9276615	32851196	HLA-DQB2	0.686	0.769	12.44	0.00042	1.53	(1.21,1.93)
rs2621377	32871089	HLA-DOB	0.508	0.612	15.75	7.20E-05	1.56	(1.26,1.93)
rs2857151	32871493	HLA-DOB	0.55	0.661	18.63	1.60E-05	1.65	(1.33,2.05)
rs9276726	32871867	HLA-DOB	0.251	0.336	12.45	0.00042	1.52	(1.21,1.9)
rs1383261	32873430	HLA-DOB	0.255	0.341	12.74	0.00036	1.52	(1.22,1.91)
rs2621332	32888296	HLA-DOB	0.55	0.66	18.45	1.70E-05	1.65	(1.33,2.05)
rs2856997	32889755	HLA-DOB	0.618	0.526	12.28	0.00046	0.66	(0.53,0.82)
rs10484565	32903011	TAP2	0.934	0.862	20.39	6.30E-06	0.44	(0.31,0.63)
rs241433	32906774	TAP2	0.558	0.626	6.72	0.0096	1.33	(1.08,1.65)
rs2281389	33167775	HLA-DPB1	0.799	0.853	7.26	0.007	1.47	(1.12,1.92)
rs3117222	33168928	HLA-DPB1	0.74	0.798	6.83	0.0089	1.39	(1.09,1.77)
rs2064478	33180245	HLA-DPB1	0.757	0.817	7.76	0.0053	1.44	(1.12,1.85)
rs3117230	33183614	HLA-DPB1	0.757	0.817	7.59	0.0059	1.43	(1.11,1.84)
rs3117016	33203495	COL11A2	0.542	0.619	8.61	0.0033	1.38	(1.12,1.7)
rs3117004	33204745	COL11A2	0.603	0.682	9.81	0.0017	1.43	(1.15,1.77)
rs3130237	33205540	COL11A2	0.817	0.875	9.1	0.0026	1.57	(1.18,2.09)
rs6901221	33206255	COL11A2	0.818	0.877	9.52	0.002	1.59	(1.19,2.13)
rs986521	33244124	COL11A2	0.768	0.701	8.16	0.0043	0.71	(0.56,0.89)
rs2855430	33249259	COL11A2	0.836	0.899	11.83	0.00058	1.75	(1.28,2.39)
rs2855425	33252352	COL11A2	0.745	0.683	6.66	0.0099	0.74	(0.59,0.93)
rs2855459	33262635	COL11A2	0.835	0.891	9.55	0.002	1.63	(1.2,2.2)

^a8.1AF_Control = allele frequency of the 8.1 haplotype associated allele in controls.^b8.1AF_Case = allele frequency of the 8.1 haplotype associated allele in cases

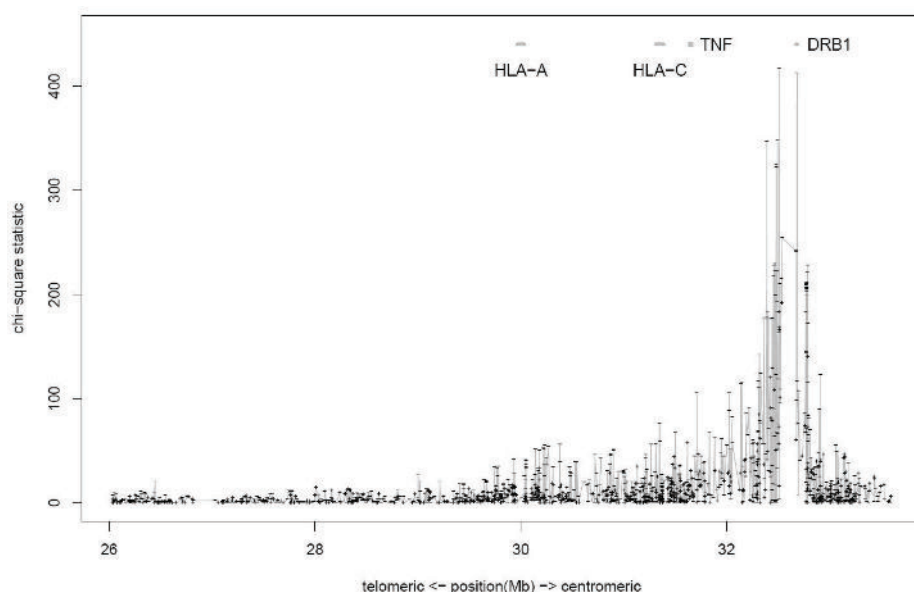


Figure 1. Chi-square test statistics from the 2-by-2 allele count table of 2,062 Case-control samples (whole dataset, 869 RA cases, 1,193 controls) of 1769 SNPs across the MHC region.

associated with RA: positively if the allele that is on the conserved 8.1 haplotype has a higher frequency in RA than in the control group, and negatively if the 8.1 allele frequency is higher in the control group. Because the complete sequence of the 8.1 haplotype is available (27), for every SNP, we were able to identify the allele that is found on the 8.1 haplotype. As an added approach, we also examined all DR3 homozygous individuals in our dataset and assessed the pattern of homozygosity across the MHC, as shown in figure 2 of Supplementary Materials. The 8.1 haplotype can be inferred from the majority homozygotes among these DR3 homozygous samples. The alleles on the 8.1 haplotype inferred this way were 100% consistent with the ones obtained from the Sanger Institute. Supplementary figure 2 shows that 12/20 DR3 homozygous individuals were homozygous over the entire haplotype extending from a region centromeric of DRB1 to the HLA-A locus, with evidence of microheterogeneity in some subjects. Thus, as expected, the 8.1 extended haplotype is quite common in our population, given its predominantly western and northern European ancestry.

Figure 2 shows single-SNP chi-square (χ^2) test statistics of 2-by-2 allele count tables for the 744 matched case control samples. (For genotype-based test re-

sults, see figure 3 of Supplementary Materials). Positive association signals with the 8.1 haplotype (*i.e.*, 8.1 allele being enriched in the RA group, labeled with +) are located in the region around HLA-C, with maximal evidence for association within the HLA-C locus. This general region was implicated in our previous studies using microsatellite typing across the MHC (16), and therefore confirms that haplotypic elements that are related to the 8.1 extended haplotype confer risk for RA, independent of the DRB1 locus. Specifically, this signal cannot be due to LD with DR3 because both cases and controls are matched 1:1 for this allele, as well as all other major DRB1 alleles. Interestingly, we also observed significant association signals in the region centromeric to the DRB1 locus in regions that include the DOB, TAP2, DPB1, and COL11A2 genes (see Figure 2 and Table 1). These signals are mainly “positive,” *i.e.*, the 8.1 allele is enriched in the RA group.

Our previous microsatellite studies of the MHC had suggested that a more

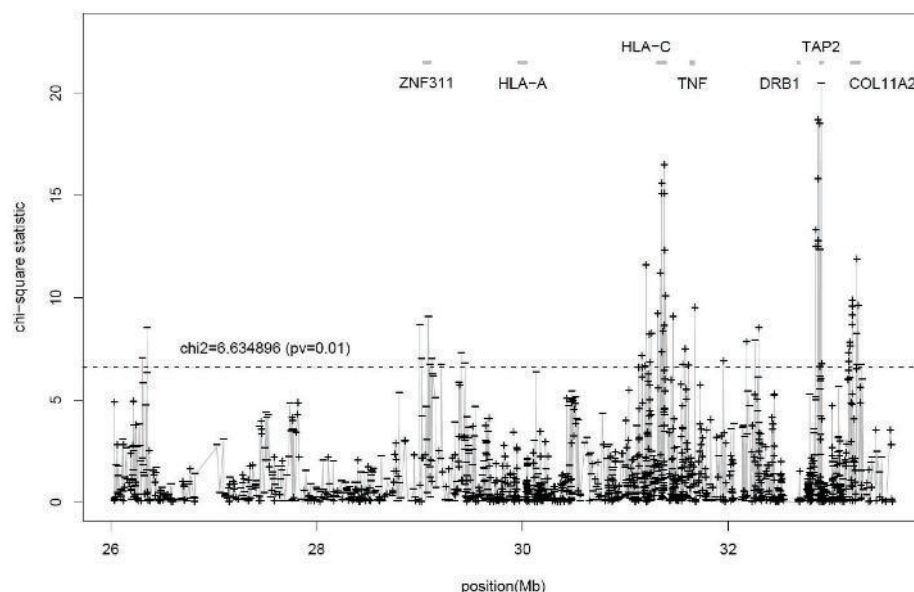


Figure 2. Chi-square test (χ^2) statistics from the 2-by-2 allele count table of 744 case-control samples matched 1:1 by DRB1 genotype (372 cases and 372 controls). Association signals are separated into two types: a) positive associations (8.1 allele frequency is greater in RA cases than controls) (+ symbol), and b) negative associations (8.1 allele frequency is greater in controls than in cases) (– symbol).

Table 2.^a Comparison of Odds Ratios among DR4 subgroups for the MHC region between ~29 and ~29.4 Mb.

Marker	position	Gene	SE(a)/X or SE(a)/SE(a) (n = 502)		0404/X but not 0401 (n = 106)		0401/X but not 0404 (n = 276)	
			OR	(95%CI)	OR	(95% CI)	OR	95% CI
rs3763338	29002291	RFP	0.58	(0.37-0.93)	0.34	(0.10-1.10)	0.55	(0.97-3.47)
rs2071788	29020287	RFP	0.63	(0.39-1.02)	0.34	(0.10-1.10)	0.54	(0.91-3.71)
rs6923005	29084051	ZNF311	0.70	(0.53-0.92)	0.69	(0.36-1.31)	0.82	(0.84-1.77)
rs6930903	29089224	ZNF311	0.70	(0.53-0.92)	0.69	(0.36-1.31)	0.82	(0.84-1.77)
rs6456889	29118804	OR2W1	0.72	(0.54-0.95)	0.72	(0.37-1.38)	0.83	(0.82-1.76)
rs9295794	29212915	OR2J3	0.62	(0.39-1.01)	0.34	(0.10-1.10)	0.57	(0.87-3.58)
rs9357078	29405646	OR5U1	0.58	(0.38-0.89)	0.18	(0.05-0.63)	0.59	(0.96-3.02)
rs9380120	29443517	OR12D3	0.60	(0.39-0.91)	0.24	(0.08-0.74)	0.59	(0.96-3.02)

^aAs in Table 1, the ORs are shown with respect to the 8.1 haplotype associated allele for each marker. Markers that show substantially increased and significant evidence of association in the DRB1*0404 subset are shown in bold.

telomeric portion of the Class I region also might be involved in RA susceptibility (16). In this case, the relevant microsatellite alleles were not related to the ancestral 8.1 haplotype, but rather could be found on certain DRB1*0404 haplotypes that were transmitted to affected subjects. It was, therefore, of interest that in the current study, evidence of additional association was observed in a region between 29.0 and 29.4 Mb, with maximal chi-square values over 8.9 for markers rs6923005 and rs6930903 (see Figure 2 and Table 1). When we examined associations in this region on the subgroup of paired sample containing any SE + DRB1*04 allele (designated SE(a) alleles for our purposes), there was increased evidence for association in this region, as shown in Table 2. The vast majority of SE(a) alleles are either DRB1*0401 or DRB1*0404 in our dataset. To determine which particular DRB1*04 allele, 0404 or 0401, is connected to this signal, we examined these associations in individuals who carried 0404 in the absence of 0401, and *vice versa*. Strikingly, even without correcting for differences in sample size, the associations in this region are overwhelmingly found in subjects carrying a DRB1*0404 allele in the absence of DRB1*0401, and not in subjects carrying DRB1*0401 in the absence of DRB1*0404 (Table 2).

For exploratory purposes, we also carried out a subgroup analysis to determine if the association signals in the Class I or centromeric Class II region are present (or absent) within any particular subsets of DRB1 genotypes, perhaps suggesting interactions with the DRB1 locus (other than LD). For this purpose we categorized genotypes into five groups, based on the designations of alleles as belonging to the SE(a), SE(b), or SE(-) groups, as described in the methods section. The rationale for these divisions relates to the generally higher odds ratios seen in the literature with SE(a) alleles compared with SE(b) alleles (12). Admittedly, these categories of risk alleles are imprecise, and were employed here only for exploratory purposes. We examined five genotypic subgroups: A) SE(-)/SE(-), B) SE(b)/SE(-), C) SE(a)/SE(-), D) SE(b)/SE(b) and SE(a)/SE(b), and E) SE(a)/SE(a). The results of these analyses showed a complex pattern of contributions to the overall chi-square (see Table 2 of Supplementary Materials). In addition to differences discussed above that reflect the contribution of 0404 subjects to the signals in Class I, there were a variety of differences among the subgroups in terms of patterns of association outside DRB1. However, these findings need to be validated in a much larger

sample set, and we wish to emphasize that none of our results are corrected for multiple testing.

DISCUSSION

In this report we have provided confirmatory evidence for the existence of additional risk genes for rheumatoid arthritis in the Class I region of the MHC that are independent of the well known risk alleles at the DRB1 locus. The data are confirmatory in the sense that our previous study using 54 microsatellite markers across the MHC (16) suggested that at least two distinct regions in the Class I region are independently associated with RA, and these regions overlap with the regions implicated in the current study. Therefore, although we have utilized over 1,700 SNP markers in the present analysis, we have not carried out strict correction for multiple testing, because there is substantial prior evidence for association in these two Class I regions. Our new data also provide evidence for additional risk alleles in the extreme centromeric portion of the Class II region, although of course these results require replication. One should also note that the study (16) compared alleles transmitted to the RA patient and alleles that are not transmitted, thus all genetic information was obtained from pedigrees with the case samples. In the current case-control study, extra control samples are used as the baseline for allele frequencies. The two studies can then be considered as independent, with the exception of the 74 NARAC cases that are included in both studies (see Methods). As a further check of our current findings, we have also run the analysis after excluding these 74 case-control pairs with results that are highly similar to those shown in Figures 1 and 2, with significant ($\chi^2 > 6.635$) associations in both Class I regions, as well in the region centromeric to DRB1 in the matched analysis (data not shown).

We have chosen an analytic approach that involves matching of cases and controls 1:1, based on DRB1 genotype. This was made possible by the large sample

sets available to us. In this way, we have completely removed the association signal at DRB1. Any residual minor mismatching at DRB1 is due to having only two digit rather than four digit typing on a proportion of SE negative alleles. To reduce error, we also confined ourselves to the evaluation of SNPs that had very high call rates (> 98%) in this region for our entire dataset. In addition, we employed a set of ancestry informative markers to reduce population stratification. Thus, we view these data as providing very strong support for the hypothesis that several regions in the MHC provide risk for RA, independent of the DRB1 locus.

Our analysis was also particularly influenced by previous data suggesting that a fragment of the conserved ancestral 8.1 haplotype carries risk alleles for RA that are independent of DRB1. This hypothesis is derived from studies in our lab (16) as well as others (15,28). As discussed below, the ancestral 8.1 haplotype has been implicated in a number of different autoimmune diseases.

For the association signal in the region around HLA-C (see Figure 2), the risk alleles are found on a common ancestral haplotype carrying the classical HLA alleles A*0101-B*0801-C*0701-DRB1*0301-DQB1*0201-DQA1*0501-DPA1*0101-DPB1*0301, commonly known as the "A1-B8-DR3" or "8.1" ancestral haplotype. The 8.1 haplotype is of considerable medical and biological interest (25). A wide array of autoimmune phenotypes has been associated with this haplotype, including systemic lupus erythematosus (29), autoimmune hepatitis, and myasthenia gravis (25). Interestingly, recent data suggest that DRB1 is not the primary locus of risk for Myasthenia Gravis (30). Truncated 8.1 haplotypes that lack the DR3 allele are associated with higher levels of acetylcholine receptor autoantibodies in subjects with Myasthenia Gravis, and the primary association with disease appears to lie telomeric to the Class II region. In addition to associations with autoimmunity, the 8.1 ancestral haplotype has also been implicated in risk for IgA deficiency and common

variable immunodeficiency (25). A low antibody response after hepatitis B immunization has also been associated with this haplotype (31), and some reports in the literature suggest that the 8.1 haplotype is associated with altered cytokine production and reduced Fc receptor function (32-34). Thus, there appears to be something immunologically distinctive about this extended haplotype.

The major association signals in RA that are related to the 8.1 haplotype encompass a number of potential candidate genes in the Class I regions, and the current data does not permit us to choose between them with any degree of confidence (see Table 1). Many significant signals are present within the HLA-C locus, and the maximal association is within the HLA-C locus itself. A substantial literature on the complex and complementary relationship between HLA-C locus polymorphisms and the extensive molecular diversity at the KIR locus is now emerging (35). Interestingly, one report has suggested that certain combinatorial relationships at these loci may predispose to rheumatoid vasculitis (36), although others have not observed such a relationship for RA in general (37). Nevertheless, complex interactions have been described for psoriatic arthritis (38), outcome of HIV infection (39), nasopharyngeal carcinoma (40), and the maternal fetal relationships underlying preeclampsia (41). The HLA-C allele found on the 8.1 haplotype is HLA-C*0701, a member of the "C-group 1" alleles that are characterized by an asparagine residue at position 80 (35). The C-group 1 alleles are ligands for inhibitory KIR2DL2/3 and activating KIR2DS2 receptors. A full characterization of HLA-C locus diversity in RA is now clearly mandated by the current findings in the context of the most up to date approaches to the analysis of the KIR locus (35).

We (16) and others (17) have proposed previously that the MHC also contains genetic risk elements outside of DRB1 that are unrelated to the extended 8.1 haplotype. In particular, we were interested in extending our prior evidence for

a signal in the Class I region found in subjects carrying DRB1*0404 haplotypes. As shown in Table 1 and Figure 2, a group of SNPs extending from 29.0-29.4 Mb are significantly associated with RA. To explore whether there is an interaction between this association and DRB1*0404, we compared the strength of this association in subjects carrying various DR4 shared epitope alleles - designated the "SE(a)" group of alleles in this discussion (see Methods) - with individuals carrying the two major DR4 risk alleles - DRB1*0401 and DRB1*0404. Strikingly, the evidence for association in the region is considerably higher in the 0404 group, despite a much lower sample size in this group (see Table 2).

Our previous study suggested that some association signals in Class I might be due to alleles that are present in a subset of *0404 haplotypes. The current study does not allow us to address this issue directly. However, regardless of the haplotypes involved, the region extending from 29.0-29.4 megabases contains risk alleles that appear to interact with the DRB1*0404 allele. In this case, the term "interact" implies that these two loci are involved in the same mechanistic pathway. It will require a much expanded sample size of cases and controls carrying DRB1*0404 to understand this finding fully.

Finally, we have developed evidence of additional independent risk effect in the region centromeric to the DRB1 locus. Areas of maximal association include the DOB, TAP2, DPB1, and COL11A2 genes. These findings are of interest because similar findings are emerging from a conditional analysis of rheumatoid arthritis in the Swedish population (Ding *et al.*, manuscript submitted).

In summary, we have shown the presence of multiple independent risk regions for RA susceptibility within the MHC complex. The data strongly suggest the need for comprehensive HLA typing of large populations of RA patients, additional SNP mapping, and resequencing of candidate genes in the

region. HLA typing should include full Class I typing so that potential interactions with KIR loci can be pursued. It is striking that after more than 30 y of investigation into the MHC relationships with rheumatoid arthritis, there is still much to be learned.

ACKNOWLEDGMENTS

Support was provided by the National Institutes of Health, RO1-AR44222 and U19 IMAGEN. The research also was supported in part by the intramural program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases, as well as the Eileen Ludwig Greenland Center for Rheumatoid Arthritis at the Feinstein Institute. LAC is supported by R01 AI065841.

REFERENCES

- Firestein GS. (2003) Evolving concepts of rheumatoid arthritis. *Nature*. 423:356-61.
- Begovich AB *et al.* (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am. J. Hum. Genet.* 75:330-7.
- Suzuki A *et al.* (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* 34:395-402.
- Plenge RM *et al.* (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis—a genome-wide study. *N. Eng. J. Med.* 357:1199-209.
- Remmers EF *et al.* (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Eng. J. Med.* 357:977-86.
- Plenge RM *et al.* (2007) Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.* 39:1477-82.
- Jawaheer D *et al.* (2003) Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multi-case families. *Arthritis Rheum.* 48:906-16.
- Cornelis F *et al.* (1998) New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc. Natl. Acad. Sci. U. S. A.* 95:10746-50.
- MacKay K *et al.* (2002) Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum.* 46:632-9.
- Stastny P. (1976) Mixed lymphocyte cultures in rheumatoid arthritis. *J. Clin. Invest.* 57:1148-57.
- Gregersen PK, Silver J, Winchester RJ. (1987) The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum.* 30:1205-13.
- Hall FC *et al.* (1996) Influence of the HLA-DRB1 locus on susceptibility and severity in rheumatoid arthritis. *Qjm.* 89:821-9.
- Gourraud PA *et al.* (2006) A new classification of HLA-DRB1 alleles differentiates predisposing and protective alleles for rheumatoid arthritis structural severity. *Arthritis Rheum.* 54:593-9.
- Mulcahy B *et al.* (1996) Genetic variability in the tumor necrosis factor-lymphotoxin region influences susceptibility to rheumatoid arthritis. *Am. J. Hum. Genet.* 59:676-83.
- Ota M *et al.* (2001) A second susceptibility gene for developing rheumatoid arthritis in the human MHC is localized within a 70-kb interval telomeric of the TNF genes in the HLA class III region. *Genomics.* 71:263-70.
- Jawaheer D *et al.* (2002) Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. *Am. J. Hum. Genet.* 71:585-94.
- Newton JL *et al.* (2004) Dissection of class III major histocompatibility complex haplotypes associated with rheumatoid arthritis. *Arthritis Rheum.* 50:2122-9.
- Jawaheer D, Lum RF, Amos CI, Gregersen PK, Criswell LA. (2004) Clustering of disease features within 512 multicase rheumatoid arthritis families. *Arthritis Rheum.* 50:736-41.
- Wolfe F, Michaud K, Gefeller O, Choi HK. (2003) Predicting mortality in patients with rheumatoid arthritis. *Arthritis Rheum.* 48:1530-42.
- Fries JF *et al.* (2002) HLA-DRB1 genotype associations in 793 white patients from a rheumatoid arthritis inception cohort: frequency, severity, and treatment bias. *Arthritis Rheum.* 46:2320-9.
- Irigoyen P *et al.* (2005) Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: Contrasting effects of HLA-DR3 and the shared epitope alleles. *Arthritis Rheum.* 52:3813-8.
- Mitchell MK, Gregersen PK, Johnson S, Parsons R, Vlahov D. (2004) The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J. Urban Health.* 81:301-10.
- Erich H *et al.* (1991) HLA-DR, DQ and DP typing using PCR amplification and immobilized probes. *Eur. J. Immunogenet.* 18:33-55.
- Tian C *et al.* (2008) Analysis and Application of European Genetic Substructure Using 300 K SNP Information. *PLoS. Genet.* 4:e4.
- Price P *et al.* (1999) The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol. Rev.* 167:257-74.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38:904-9.
- Traherne JA *et al.* (2006) Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS. Genet.* 2:e9.
- Moxley G, Cohen HJ. (2002) Genetic studies, clinical heterogeneity, and disease outcome studies in rheumatoid arthritis. *Rheum. Dis. Clin. North Am.* 28:39-58.
- Graham RR *et al.* (2002) Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am. J. Hum. Genet.* 71:543-53.
- Vandiedonck C *et al.* (2004) Pleiotropic effects of the 8.1 HLA haplotype in patients with autoimmune myasthenia gravis and thymus hyperplasia. *Proc. Natl. Acad. Sci. U. S. A.* 101:15464-9.
- Alper CA *et al.* (1989) Genetic prediction of non-response to hepatitis B vaccine. *N. Engl. J. Med.* 321:708-12.
- Caruso C *et al.* (1997) HLA-B8,DR3 haplotype affects lymphocyte blood levels. *Immunol. Invest.* 26:333-40.
- Caruso C, Bellina L. (1983) B8, DR3 antigens and production of human leukocyte migration inhibitory factor (LIF) by mononuclear cells stimulated with concanavalin A (Con A). *Tissue Antigens.* 22:167-9.
- Candore G, Cigna D, Gervasi F, Colucci AT, Modica MA, Caruso C. (1994) *In vitro* cytokine production by HLA-B8,DR3 positive subjects. *Autoimmunity.* 18:121-32.
- Single RM *et al.* (2007) Global diversity and evidence for coevolution of KIR and HLA. *Nat. Genet.* 39:1114-9.
- Yen JH *et al.* (2001) Major histocompatibility complex class I-recognizing receptors are disease risk genes in rheumatoid arthritis. *J. Exp. Med.* 193:1159-67.
- Middleton D, Meenagh A, Wright GD. (2007) No association in frequency of KIR receptors in patients with rheumatoid arthritis from Northern Ireland. *Tissue Antigens.* 69:577-82.
- Martin MP *et al.* (2002) Cutting edge: susceptibility to psoriatic arthritis: influence of activating killer Ig-like receptor genes in the absence of specific HLA-C alleles. *J. Immunol.* 169:2818-22.
- Martin MP *et al.* (2007) Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nat. Genet.* 39:733-40.
- Butsch Kovacic M *et al.* (2005) Variation of the killer cell immunoglobulin-like receptors and HLA-C genes in nasopharyngeal carcinoma. *Cancer Epidemiol. Biomarkers Prev.* 14:2673-7.
- Hiby SE *et al.* (2004) Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. *J. Exp. Med.* 200:957-65.