ARTHRITIS & RHEUMATISM
Vol. 60, No. 1, January 2009, pp 47–52
DOI 10.1002/art.24180

The Chromosome 7q Region Association With Rheumatoid Arthritis in Females in a British Population Is Not Replicated in a North American Case–Control Series

Benjamin D. Korman,[1] Michael F. Seldin,[2] Kimberly E. Taylor,[3] Julie M. Le,[1] Annette T. Lee,[4] Robert M. Plenge,[5] Christopher I. Amos,[6] Lindsey A. Criswell,[3] Peter K. Gregersen,[4] Daniel L. Kastner,[1] and Elaine F. Remmers[1]

Objective. **The single-nucleotide polymorphism (SNP) rs11761231 on chromosome 7q has been reported to be sexually dimorphic marker for rheumatoid arthritis (RA) susceptibility in a British population. We sought to replicate this finding and to better characterize susceptibility alleles in the region in a North American population.**

Methods. **DNA from 2 North American collections of RA patients and controls (1,605 cases and 2,640 controls) was genotyped for rs11761231 and 16 additional chromosome 7q tag SNPs using Sequenom iPlex assays. Association tests were performed for each collection and also separately, contrasting male cases with male controls and female cases with female controls. Principal components analysis (EigenStrat) was used to determine association with RA before and after adjusting for population stratification in the subset of the samples for which there were whole-genome SNP data (772 cases and 1,213 controls).**

Results. **We failed to replicate an association of the 7q region with RA. Initially, rs11761231 showed evidence for association with RA in the North American Rheumatoid Arthritis Consortium (NARAC) collection ($P = 0.0073$), and rs11765576 showed association with RA in both the NARAC ($P = 0.038$) and RA replication ($P = 0.0013$) collections. These markers also exhibited sex differentiation. However, in the whole-genome subset, neither SNP showed significant association with RA after correction for population stratification.**

Conclusion. **While 2 SNPs on chromosome 7q appeared to be associated with RA in a North American cohort, the significance of this finding did not withstand correction for population substructure. Our results emphasize the need to carefully account for population structure to avoid false-positive disease associations.**

It has long been recognized that sex is a major risk factor for the development of autoimmune disease and that females are at an increased risk for these conditions. In the case of rheumatoid arthritis (RA), the

Supported by the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH. Mr. Korman's work was supported by the NIH Clinical Research Training Program, a public–private partnership between the Foundation for the NIH and Pfizer, Inc. Dr. Plenge's work was supported by the NIH (grant K08-AI-55314-3), the Research and Education Foundation of the American College of Rheumatology, the Burroughs Wellcome Fund (Career Awards for Medical Scientists), and the William Randolph Hearst Fund of Harvard University. Dr. Amos' work was supported by the NIH (grant R01-AR-44422). Dr. Criswell's work was supported by the NIH (grants R01-AI-065841, K24-AR-02175, and N01-AR-72232) and by the Rosalind Russell Medical Research Center for Arthritis. Dr. Gregersen's work was supported by the NIH (grants R01-AR-44422 and N01-AI-95386). The studies were funded by the National Center for Research Resources, USPHS, and carried out in part at the General Clinical Research Center, Moffitt Hospital, University of California, San Francisco (grant 5-M01-RR-00079) and at the General Clinical Research Center, Feinstein Institute for Medical Research (grant M01-RR-018535).

[1]Benjamin D. Korman, BS, Julie M. Le, BS, Daniel L. Kastner, MD, PhD, Elaine F. Remmers, PhD: National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, Maryland; [2]Michael F. Seldin, MD, PhD: University of California, Davis; [3]Kimberly E. Taylor, PhD, MPH, Lindsey A. Criswell, MD, MPH: University of California, San Francisco; [4]Annette T. Lee, PhD, Peter K. Gregersen, MD: Feinstein Institute for Medical Research, Manhasset, New York; [5]Robert M. Plenge, MD, PhD: Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, and Broad Institute of MIT and Harvard, Cambridge, Massachusetts; [6]Christopher I. Amos, PhD: University of Texas, and M. D. Anderson Cancer Center, Houston, Texas.

Dr. Plenge has received consulting fees, speaking fees, and/or honoraria from Genentech and Biogen Idec (less than $10,000 each).

Address correspondence and reprint requests to Elaine F. Remmers, PhD, National Institutes of Health, 10 Center Drive, 10/10C101, MSC 1849, Bethesda, MD 20892-1849. E-mail: remmerse@mail.nih.gov.

Submitted for publication January 10, 2008; accepted in revised form September 25, 2008.

female:male ratio is ~3:1 (1). Furthermore, RA disease characteristics tend to be different in males and females (2). Despite the suspicion that genetic factors underlie these observations, the genes that have been implicated in autoimmunity have yet to explain the female predominance of RA or any other autoimmune disorder.

Genetic studies have identified a number of loci that are associated with RA susceptibility, including *HLA–DRB1* (3), *PTPN22* (4), and, more recently, *STAT4* (5), *TRAF1/C5* (6), and a region near *TNFAIP3* on chromosome 6q23 (7,8). Among the findings in their whole-genome association study of 7 diseases (9), the Wellcome Trust Case Control Consortium (WTCCC) identified a single-nucleotide polymorphism (SNP), rs11761231, on chromosome 7q, which exhibited sexual dimorphism, showing a statistically significant association with RA susceptibility only in females ($P = 6.8 \times 10^{-8}$ in females and $P = 0.68$ in males). The authors suggested that this variant in RA might represent one of the first sex-differentiated genetic effects in human autoimmune disease.

In the present study, we sought to confirm the association of rs11761231 with RA in females by genotyping North American RA patients and healthy controls. Furthermore, to better characterize the region surrounding this SNP, we also genotyped 16 additional SNPs to determine whether other nearby markers may have similar or stronger disease association.

PATIENTS AND METHODS

Subjects. DNA from North American RA patients and unrelated control subjects, both groups of European ancestry, was obtained from 2 previously reported case–control collections, the North American Rheumatoid Arthritis Consortium (NARAC) series and the RA replication series (5). The NARAC cases included 1 affected member from each family of European descent from the NARAC collection of affected sibling pairs collected throughout North America (10). The replication series cases were self-described Caucasians whose DNA was obtained through the National Data Bank for Rheumatic Diseases (Wichita, KS) (11), the National Inception Cohort of Rheumatoid Arthritis Patients (nationwide US) (12), and the Study of New-Onset Rheumatoid Arthritis (North America) (13).

All of the cases met the American College of Rheumatology (formerly, the American Rheumatism Association) 1987 revised criteria for the classification of RA (14). Of the NARAC cases, 100% had longstanding disease, 81.7% were positive for rheumatoid factor, 80.5% were positive for anti–cyclic citrullinated peptide (anti-CCP), 80.9% were positive for the shared epitope, 80.1% were female, and 94.2% had hand erosions on radiographs read by a single radiologist. Of the replication series cases, 72.2% were female, 98.5% were positive for anti-CCP, 48.5% had longstanding disease, and the group had a mean age at onset of 49.7 ± 14.1 years. The controls for these collections were population controls, not evaluated for disease, from the New York Cancer Project (15). DNA for these controls was collected from donors from New York City and the surrounding area. The controls selected were between the ages of 30 and 60 years and Caucasian by self report, and where possible, they were matched to the cases by self-reported country of origin of their grandparents and by decade of birth for age. In total, the samples analyzed included DNA from 1,605 independent RA cases and 2,640 independent population controls of European ancestry. We also performed a subset analysis limited to DNA samples (from 772 cases and 1,213 controls) for which whole-genome genotype data were available (6). Informed consent was obtained from every subject, and approval of the local Institutional Review Board was secured at every recruitment site prior to the start of enrollment.

Selection of SNPs. In addition to rs11761231, we used the tag SNP Picker utility available at the International HapMap Consortium Web site (www.hapmap.org), which uses the Tagger algorithm (16), to select additional tag SNPs. These 19 tag SNPs capture, with pairwise $r^2 > 0.8$, all HapMap variants with a >5% minor allele frequency located in close proximity to and within the expressed sequence tag (EST) DA600502, which contains the reported SNP, rs11761231.

Genotyping. Multiplex SNP assays were designed using Sequenom RealSNP software (www.realsnp.com); 2 SNPs, rs2909480 and rs12536699, failed to produce genotypes in the designed assays. The remaining 17 SNPs were successfully genotyped by the iPlex Gold protocol and the genotypes determined by SpectroTyper software (Sequenom, San Diego, CA). Calls were evaluated and edited by cluster analysis performed with the SpectroTyper software. Deidentified cases and controls were analyzed together. Genotyping accuracy was evaluated by 2 methods. In a set of 154 controls genotyped for 17 SNPs in duplicate (duplicates on different assay plates), the genotype concordance rate was 99.9%. Additionally, 4 of the 7q region SNPs were genotyped in the recently reported whole-genome association study (6). A concordance rate of 98.7% was found in the 772 whole-genome association study cases and 1,213 controls genotyped on both the Sequenom and Illumina platforms.

Statistical analysis. After genotyping, SNP markers were evaluated for significant deviation from Hardy-Weinberg equilibrium or low minor allele frequencies. We planned to exclude markers with disequilibrium P values of < 0.005 in controls or minor allele frequency <0.01 to avoid errors in genotyping or insufficient power, but all markers genotyped passed these measures. SNPs were then analyzed for association by comparison of the minor allele frequency in cases and controls, with significance determined by a chi-square test; this was also done separately for males and females. Linkage disequilibrium patterns in the 7q region were determined using Haploview version 4.0 software (17). A principal components–based method, EigenStrat (18), was used to adjust for population structure. The EigenStrat analysis did not include the following intervals due to strong linkage disequilibrium: chromosome 6 (24–36 Mb), chromosome 8 (8–12 Mb), and chromosome 17 (40–43 Mb). The genomic control inflation factor (λ_{gc}) stabilized at 1.06 at principal component 6, and all

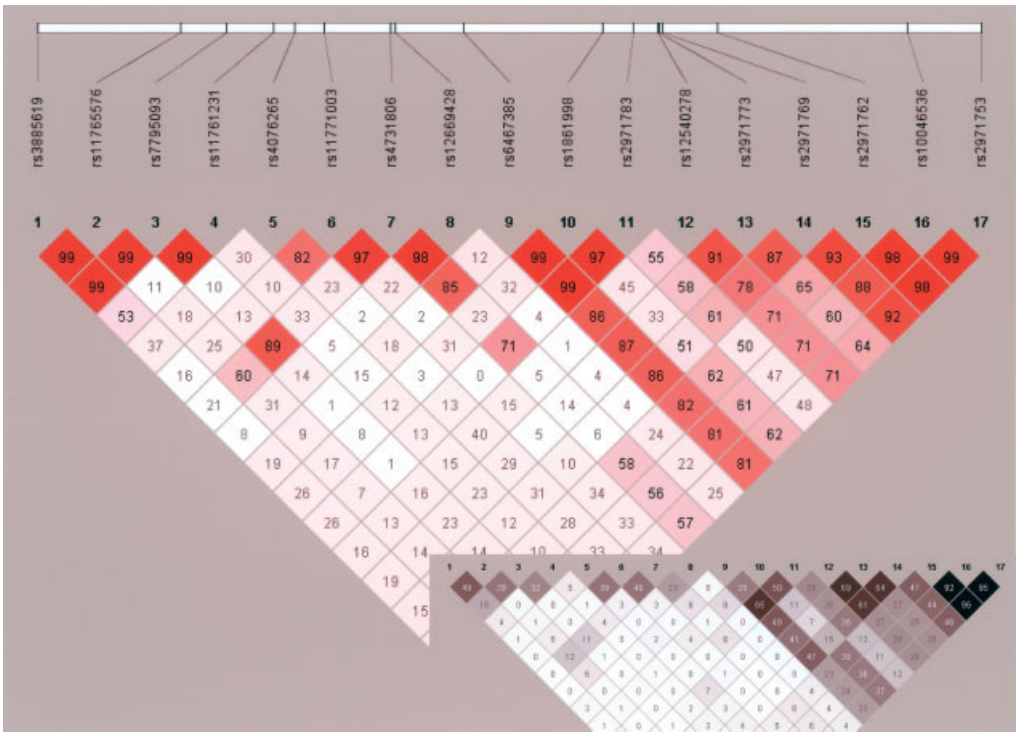

Figure 1. Linkage disequilibrium of the genotyped region of chromosome 7q32.3. The larger map represents D′; the smaller inset map represents r^2. All genotyped single-nucleotide polymorphisms are shown in the order in which they appear on the chromosome.

adjusted chi-square tests are shown for principal component 6 corrected for the residual λ_{gc}.

RESULTS

Chromosome 7q linkage disequilibrium structure. Despite the close proximity of the 17 SNPs genotyped, linkage disequilibrium in the 908-kb region was poor. The marker rs11761231 had a D′ value of >0.9 with only 1 of the other 16 markers (rs7795093); however, the r^2 value was only 0.32 (see Figure 1).

SNP association with RA in the NARAC and RA replication collections. Of the 17 SNPs genotyped, 2 showed evidence of association with RA. The first, rs11761231, replicated the WTCCC findings of a significant association with RA and sexual dimorphism in the NARAC collection (see Table 1). However, this SNP was not associated with disease in the RA replication collection. Another marker, rs11765576, showed evidence for association with RA in both the NARAC and RA replication collections and also showed sex differentiation. These markers were independent, with D′ = 0.11 and $r^2 = 0$. In comparing the 2 collections, we noted that the minor allele frequencies, especially for rs11761231, were quite different between the control groups, suggesting that population stratification might be contributing to these results (see Table 1).

SNP association with RA after correction for population structure. In order to determine the effect of population structure on our results, we analyzed the subset of samples for which whole-genome SNP data were available, both before and after correction for stratification using EigenStrat (18). Before correction for stratification, this subset showed significant evidence of association with RA and a sex-differentiated effect for both rs11761231 and rs11765576. However, after accounting for population differences by correcting the association results using principal components analysis, none of the associations remained statistically significant (see Table 2).

DISCUSSION

This study demonstrates some of the difficulties associated with genetic disease association studies. The 7q region identified by the WTCCC is an intriguing region for a number of reasons (9). Even without sex

Table 1. Sex-differentiated analysis of 2 RA case–control series*

SNP, sample collection, sex	No. of cases/ no. of controls	Allele frequency		χ^2	P
		Cases	Controls		
rs11761231					
NARAC					
M + F	607/1,315	0.315	0.361	7.19	0.0073
M	120/329	0.338	0.347	0.06	0.8022
F	486/986	0.309	0.366	8.57	0.0034
RA replication					
M + F	998/1,325	0.364	0.338	3.15	0.0760
M	275/613	0.386	0.335	4.05	0.0442
F	722/677	0.356	0.343	0.49	0.4849
Pooled					
M + F	1,605/2,640	0.346	0.350	0.13	0.7182
M	395/942	0.371	0.339	2.36	0.1247
F	1,208/1,663	0.337	0.356	2.16	0.1418
rs11765576					
NARAC					
M + F	607/1,315	0.410	0.374	4.32	0.0378
M	120/329	0.397	0.392	0.03	0.8739
F	486/986	0.414	0.368	5.51	0.0189
RA replication					
M + F	998/1,325	0.409	0.362	10.37	0.0013
M	275/613	0.405	0.368	2.11	0.1463
F	722/677	0.411	0.356	8.56	0.0034
Pooled					
M + F	1,605/2,640	0.410	0.368	13.97	0.0002
M	395/942	0.402	0.376	1.58	0.2092
F	1,208/1,663	0.412	0.363	13.55	0.0002

* Shown are minor allele frequencies and association test results for 2 chromosome 7q single-nucleotide polymorphisms (SNPs) genotyped among males and females. These data comprise the complete case–control collections previously described (5). RA = rheumatoid arthritis; NARAC = North American Rheumatoid Arthritis Consortium.

differentiation, the rs11761231 *P* value in the WTCCC study suggested an association with RA. When sex was considered, the fact that the association with this SNP was strong in females and not detected in males constituted even more intriguing evidence for this variant in RA, a female-predominant disease. These data, coupled with our observation that this SNP is located within a novel EST derived from human RA synoviocytes, motivated us to attempt to replicate the association and to evaluate additional variants from this genomic region.

Our genotyping of North American RA cases and controls initially gave supportive evidence for association of rs11761231, the SNP identified by the WTCCC as a marker for a sexually dimorphic risk factor

Table 2. RA-associated allele frequencies and association test results for 2 chromosome 7q SNPs genotyped among the subset of samples with whole-genome SNP data analyzed with and without EigenStrat adjustment for population stratification*

SNP, sex	No. of cases/ no. of controls	Allele frequency		OR (95% CI)	Unadjusted		Adjusted for population substructure	
		Cases	Controls		χ^2	P	χ^2	P
rs11761231								
M + F	772/1,213	0.325	0.361	1.17 (1.02–1.35)	4.33	0.038	2.42	0.12
M	92/293	0.331	0.349	1.08 (0.76–1.56)	0.49	0.49	0.094	0.78
F	680/920	0.324	0.365	1.20 (1.03–1.39)	4.46	0.035	2.76	0.097
rs11765576								
M + F	772/1,213	0.414	0.373	1.18 (1.04–1.35)	6.32	0.012	0.64	0.42
M	92/293	0.421	0.394	1.12 (0.80–1.58)	0.25	0.62	0.12	0.73
F	680/920	0.413	0.367	1.21 (1.05–1.41)	6.87	0.0088	1.76	0.18

* OR = odds ratio; 95% CI = 95% confidence interval (see Table 1 for other definitions).

for RA. The evidence was strongest in females of the NARAC case–control collection. Furthermore, we identified a second SNP, rs11765576, which also had stronger evidence for association in females than in males.

Interestingly, the minor allele frequencies of some of the SNPs in the region varied between the collections, even between the 2 control groups (see Table 1). We therefore sought to determine whether population admixture or population substructure differences affected the associations we observed. By limiting our analysis to the roughly one-half of individuals from the NARAC and RA replication cohorts who had also been genotyped as part of the NARAC/Epidemiological Investigation of Rheumatoid Arthritis whole-genome association scan (6) and using EigenStrat software, we attempted to identify whether such unseen nuances existed in our case and control population structures. Six principal components were necessary to control for substructure in this data set. For these analyses, there was no further dropoff in the chi-square test statistic for any of the SNPs examined after principal component 6 (principal components 7–10). Most of the dropoff was in principal components 1 and 2, suggesting that this was at least partly due to European ancestry north/south differences (19).

Correction for population substructure is greatest for markers in which allele frequency differences correlate with one or more of the principal components identified by the genome-wide data. Markers without these subpopulation allele frequency differences are relatively unaffected by these corrections. The RA-associated *STAT4* SNP, rs7574865, is an example of a SNP for which the correction for population substructure does not reduce the evidence for association (5). We found, however, that after accounting for population stratification using principal components, the associations between RA and SNPs on 7q32 were no longer significant in our population.

In this study, RA cases and population controls were derived from the genetically diverse Caucasian North American population. The cases were recruited from centers across North America, whereas the controls were recruited only from New York. However, because both the case and control populations are derived from individuals of diverse European genetic backgrounds, it is unlikely that matching for geographic location would substantially reduce the genetic diversity or the chance for stratification. Our results add to mounting evidence that in genetic association studies in North American and other genetically complex populations, stratification should be expected, even when rigorous methods are used to match cases and controls. To properly interpret association results in these populations, it is therefore necessary to apply methods that detect and correct for the stratification.

In contrast to the current study, after recent non-European migrants were excluded from the recent WTCCC study, in which the 7q SNP association was originally identified (9), there was little evidence for substructure as measured by extreme differences in SNP allele frequency in individuals from different geographic areas within the UK. Indeed, the uncorrected λ_{gc} in the British RA study was 1.03, compared with 1.43 in the NARAC genome-wide association study, indicating that the WTCCC study had a much more genetically homogeneous population. These observations led the investigators to conclude that it was unnecessary to correct the associations for stratification in their study. Although the allele frequency of this SNP did not exhibit extreme geographic variation ($P < 10^{-6}$), it would be interesting to determine whether a stratification correction, such as the principle components–based correction used here, would nonetheless influence this particular association. Interestingly, the association of the WTCCC-identified SNP, rs11761231, with RA has recently failed to be replicated in a British RA replication collection (7).

The current study raises the issue of population stratification effects in case–control studies, particularly in complex populations. Because sample sizes used are often very large and the allele frequency differences detected are modest, even minor differences in the racial or ethnic makeup of cases and controls have the potential to create false-positive associations reflective of these differences rather than disease-associated genetic differences between cases with disease and healthy controls. This observation helps explain the inherent difficulty in replicating candidate gene association studies performed in complex populations. It is not always easy to implement stratification correction given the requirement for whole-genome association data or within-European-ancestry informative marker (20,21) genotypes for each individual. However, controlling for stratification should not only avoid false-positive associations, it should also increase the power to detect true associations.

AUTHOR CONTRIBUTIONS

Dr. Remmers had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study design. Korman, Seldin, Lee, Amos, Criswell, Gregersen, Kastner, Remmers.

Acquisition of data. Korman, Le, Criswell, Kastner, Remmers.
Analysis and interpretation of data. Korman, Seldin, Taylor, Lee, Plenge, Amos, Criswell, Gregersen, Kastner, Remmers.
Manuscript preparation. Korman, Seldin, Taylor, Le, Lee, Plenge, Amos, Criswell, Gregersen, Kastner, Remmers.
Statistical analysis. Korman, Seldin, Amos, Criswell, Remmers.

REFERENCES

1. Silman AJ, Pearson JE. Epidemiology and genetics of rheumatoid arthritis. Arthritis Res Ther 2002;4 Suppl 3:S265–72.
2. Laivoranta-Nyman S, Luukkainen R, Hakala M, Hannonen P, Mottonen T, Yli-Kerttula U, et al. Differences between female and male patients with familial rheumatoid arthritis. Ann Rheum Dis 2001;60:413–5.
3. Newton JL, Harney SM, Wordsworth BP, Brown MA. A review of the MHC genetics of rheumatoid arthritis. Genes Immun 2004;5: 151–7.
4. Gregersen PK, Lee HS, Batliwalla F, Begovich AB. PTPN22: setting thresholds for autoimmunity. Semin Immunol 2006;18: 214–23.
5. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357:977–86.
6. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, et al. TRAF1–C5 as a risk locus for rheumatoid arthritis—a genomewide study. N Engl J Med 2007;357:1199–209.
7. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis association at 6q23. Nat Genet 2007;39: 1431–3.
8. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 2007;39:1477–82.
9. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78.
10. Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Etzel C, et al, for the North American Rheumatoid Arthritis Consortium. Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. Arthritis Rheum 2003;48:906–16.
11. Wolfe F, Michaud K, Gefeller O, Choi HK. Predicting mortality in patients with rheumatoid arthritis. Arthritis Rheum 2003;48: 1530–42.
12. Fries JF, Wolfe F, Apple R, Erlich H, Bugawan T, Holmes T, et al. HLA–DRB1 genotype associations in 793 white patients from a rheumatoid arthritis inception cohort: frequency, severity, and treatment bias. Arthritis Rheum 2002;46:2320–9.
13. Irigoyen P, Lee AT, Wener MH, Li W, Kern M, Batliwalla F, et al. Regulation of anti–cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA–DR3 and the shared epitope alleles. Arthritis Rheum 2005;58:3813–8.
14. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
15. Mitchell M, Gregersen P, Johnson S, Parsons R, Vlahov D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. J Urban Health 2004;81:301–10.
16. De Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217–23.
17. Barrett J, Fry B, Maller J, Daly M. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21: 263–5.
18. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.
19. Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, Silva G, et al. European population substructure: clustering of northern and southern populations. PLoS Genet 2006;2:e143.
20. Seldin MF, Price AL. Application of ancestry informative markers to association studies in European Americans. PLoS Genet 2008; 4:e5.
21. Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C, et al. Analysis and application of European genetic substructure using 300 K SNP information. PLoS Genet 2008;4:e4.