Please cite this article in press as: Kurreeman et al., Use of a Multiethnic Approach to Identify Rheumatoid- Arthritis-Susceptibility Loci, 1p36 and 17q12, The American Journal of Human Genetics (2012), doi:10.1016/j.ajhg.2012.01.010

REPORT

Use of a Multiethnic Approach to Identify Rheumatoid-Arthritis-Susceptibility Loci, 1p36 and 17q12

Fina A.S. Kurreeman,[1,2,4,5] Eli A. Stahl,[1,2,4] Yukinori Okada,[6,7] Katherine Liao,[1,2] Dorothée Diogo,[1,2,4] Soumya Raychaudhuri,[1,2,4] Jan Freudenberg,[8] Yuta Kochi,[6,7] Nikolaos A. Patsopoulos,[2,3,4] Namrata Gupta,[4] CLEAR investigators,[16] Cynthia Sandor,[1,2,4] So-Young Bang,[9] Hye-Soon Lee,[9] Leonid Padyukov,[10] Akari Suzuki,[6] Kathy Siminovitch,[11,12,13] Jane Worthington,[14] Peter K. Gregersen,[8] Laura B. Hughes,[15] Richard J. Reynolds,[15] S. Louis Bridges, Jr.,[15] Sang-Cheol Bae,[9] Kazuhiko Yamamoto,[6,7] and Robert M. Plenge[1,2,4,*]

We have previously shown that rheumatoid arthritis (RA) risk alleles overlap between different ethnic groups. Here, we utilize a multiethnic approach to show that we can effectively discover RA risk alleles. Thirteen putatively associated SNPs that had not yet exceeded genome-wide significance ($p < 5 \times 10^{-8}$) in our previous RA genome-wide association study (GWAS) were analyzed in independent sample sets consisting of 4,366 cases and 17,765 controls of European, African American, and East Asian ancestry. Additionally, we conducted an overall association test across all 65,833 samples (a GWAS meta-analysis plus the replication samples). Of the 13 SNPs investigated, four were significantly below the study-wide Bonferroni corrected p value threshold ($p < 0.0038$) in the replication samples. Two SNPs (rs3890745 at the 1p36 locus [$p = 2.3 \times 10^{-12}$] and rs2872507 at the 17q12 locus [$p = 1.7 \times 10^{-9}$]) surpassed genome-wide significance in all 16,659 RA cases and 49,174 controls combined. We used available GWAS data to fine map these two loci in Europeans and East Asians, and we found that the same allele conferred risk in both ethnic groups. A series of bioinformatic analyses identified *TNFRSF14-MMEL1* at the 1p36 locus and *IKZF3-ORMDL3-GSDMB* at the 17q12 locus as the genes most likely associated with RA. These findings demonstrate empirically that a multiethnic approach is an effective strategy for discovering RA risk loci, and they suggest that combining GWASs across ethnic groups represents an efficient strategy for gaining statistical power.

Rheumatoid arthritis (RA [MIM 180300]) is characterized by chronic inflammation and destruction of the synovial joints, the latter of which leads to progressive damage and disability. Both environmental and genetic factors are involved in the etiology of RA, and there is an estimated genetic component between 50% and 60%.[1] Candidate-gene studies and genome-wide association studies (GWASs) have begun to unravel the complex genetic architecture of RA, for which >35 genetic loci have been identified to date.[2–6] All of these loci have met a stringent genome-wide significance threshold ($p < 5 \times 10^{-8}$). However, the SNPs identified individually increase risk in small increments and collectively only explain ~18% of the overall genetic heritability.[2] One of the explanations for this "missing heritability" is that many more common risk alleles remain to be discovered in larger studies.

The majority of RA risk alleles have been identified and validated in patients who are of European ancestry and are seropositive for disease-specific autoantibodies (either anticitrullinated protein antibodies [ACPAs] or rheumatoid factor [RF]). Several studies have shown that validated RA risk alleles contribute to risk in other ethnic groups, including patients of African, Hispanic, and Asian ancestry.[7–10] Although two GWASs have been performed in individuals of East Asian ancestry (Japanese and Korean),[7,9] no study has integrated data in a multiethnic fashion to discover new RA risk loci. Indeed, there has been debate in the literature as to whether common variants discovered by GWASs will be ethnic specific or contribute to risk across all ethnic groups, not only in RA but also more generally in other complex traits.[11] As proof of concept that integrating multiethnic data is an effective strategy for discovering RA risk loci, we used a multiethnic replication panel to test 13 SNPs that did not reach genome-wide significance ($p < 5 \times 10^{-8}$) in our previous GWAS.

The initial GWAS meta-analysis used as a starting point for the current study has been described in greater detail elsewhere.[2] In brief, the GWAS consisted of 5,505 RA cases and 22,603 controls, and follow-up SNP genotyping was performed in a replication collection of 6,768 cases and 8,806 controls of European ancestry (Table S1, available

[1]Division of Rheumatology, Immunology, and Allergy, [2]Division of Genetics, [3]Division of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA; [4]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, 02142, USA; [5]Department of Rheumatology, Leiden University Medical Center, Leiden 2333ZA, Netherlands; [6]Laboratory for Autoimmune Diseases and Statistical genetics, Center for Genomic Medicine, RIKEN, Yokohama, Japan; [7]Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; [8]Feinstein Institute for Medical Research and North Shore-Long Island Jewish Health System, Manhasset, New York, NY, USA; [9]Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, South Korea; [10]Rheumatology Unit, Department of Medicine, Karolinska Institutet, Stockholm 17176, Sweden; [11]Department of Medicine, Immunology, Molecular Genetics, University of Toronto, Toronto, ON, Canada; [12]Mount Sinai Hospital Samuel Lunenfeld Research Institute, Toronto, ON, Canada; [13]Toronto General Research Institute, Toronto, ON, Canada; [14]Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; [15]Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL, USA
[16]See Acknowledgments for a list of CLEAR investigators
*Correspondence: rplenge@partners.org
DOI 10.1016/j.ajhg.2012.01.010.

Please cite this article in press as: Kurreeman et al., Use of a Multiethnic Approach to Identify Rheumatoid- Arthritis-Susceptibility Loci, 1p36 and 17q12, The American Journal of Human Genetics (2012), doi:10.1016/j.ajhg.2012.01.010

Table 1. Previous Association of 13 SNPs in RA GWAS Meta-Analysis and Replication Sample Sets of European Ancestry

SNP Characteristics[a]				Alleles[b]			GWAS[c]		Previous Replication[d]		Previous Joint Meta-Analysis[e]		
SNP	Chr.	Position (bp)	Genes	A1	A2	Risk	OR	p	OR	$p_{one\text{-}tailed}$	OR	p_{Joint}	Q
rs3890745*	1p36	2,543,484	*TNFRSF14, MMEL1*	T	C	T	1.13	1.14×10^{-6}	1.08	0.069	1.12	5.28×10^{-7}	0.23
rs7543174	1q21	152,794,296	*IL6R*	C	T	C	1.14	6.00×10^{-5}	1.07	0.012	1.10	1.01×10^{-5}	0.10
rs12746613*	1q23	159,733,666	*FCGR2A*	T	C	T	1.13	4.84×10^{-4}	1.10	0.011	1.12	3.34×10^{-5}	0.19
rs10919563*	1q32	196,967,065	*PTPRC*	G	A	G	1.15	1.80×10^{-4}	1.11	0.005	1.13	6.27×10^{-6}	0.87
rs13119723	4q27	123,437,763	*IL2, IL21*	G	A	A	0.89	1.03×10^{-3}	0.87	6.75×10^{-5}	0.88	5.48×10^{-7}	0.46
rs11594656	10p15	6,162,015	*IL2RA*	T	A	T	1.10	8.96×10^{-4}	1.05	0.037	1.08	0.0003	0.90
rs2793108	10p11	31,419,111	*ZEB1*	T	C	T	1.07	5.01×10^{-3}	1.08	0.001	1.07	3.98×10^{-5}	0.82
rs540386*	11p12	36,481,869	*TRAF6*	C	T	C	1.14	1.56×10^{-4}	1.09	0.013	1.11	1.69×10^{-5}	0.09
rs3184504	12q24	110,368,991	*SH2B3*	T	C	T	1.07	2.81×10^{-3}	1.09	2.48×10^{-4}	1.08	4.68×10^{-6}	0.06
rs8045689	16p11	28,895,770	*CD19, NFATC2IP*	C	T	C	1.14	5.35×10^{-5}	1.06	0.014	1.09	2.45×10^{-5}	0.35
rs2872507	17q12	35,294,289	*IKZF3, ORMDL3, GSDMB*	G	A	A	0.92	1.66×10^{-4}	0.93	0.002	0.92	2.87×10^{-6}	0.83
rs11203203	21q22	42,709,255	*UBASH3A*	A	G	A	1.11	3.14×10^{-5}	1.07	0.017	1.09	4.64×10^{-6}	0.49
rs5754217	22q11	20,269,675	*UBE2L3*	T	G	T	1.10	1.24×10^{-3}	1.07	0.010	1.09	7.98×10^{-5}	0.96

We list results of 13 SNPs that were previously suggestive in two GWAS meta-analysis studies but that had not yet reached the conservative genome-wide significance level at $p < 5 \times 10^{-8}$. All odds ratios (ORs) are reported with respect to allele A1. The following abbreviations are used: Chr., chromosomal region; and Q, Cochran's Q test for heterogeneity of odds ratios across all sample collections.
[a]SNPs marked by an asterisk were suggestive in Raychaudhuri et al.,[6] and all remaining SNPs were suggestive in Stahl et al.[2] Chromosome and base-pair positions are with respect to the UCSC HG18 build.
[b]Alleles at each SNP (A1 and A2) and the allele that was previously identified as the risk allele (Risk).
[c]Results from a recently performed GWAS meta-analysis incorporating principal components for the correction of population stratification.[4]
[d]Results from previous replication efforts as part of Raychaudhuri et al.[6] and Stahl et al.[2]
[e]Results of the meta-analysis (performed with the inverse-variance method of meta-analysis) of all available datasets.[17]

online). Table 1 displays a summary of the SNPs from this study and suggestive evidence of association, which we defined as SNPs for which $p_{GWAS} \leq 0.005$ and $p_{replication} \leq 0.05$ but which did not reach $p_{overall} < 5 \times 10^{-8}$ in all samples combined. We included one SNP, rs3890745 at the 1p36 locus, for which $p_{replication} = 0.069$ from this study because it had previously demonstrated suggestive evidence of replication,[6] but it had not yet reached $p_{overall} < 5 \times 10^{-8}$ in all samples combined. The purpose of the current study is to test these 13 putatively RA-associated SNPs in an independent, multiethnic collection of RA case-control samples and provide definitive evidence ($p_{overall} < 5 \times 10^{-8}$) of association with RA risk.

For the multiethnic replication study described herein, we used an independent collection of 4,366 RA cases and 17,765 controls (Table 2). All RA patients satisfied international criteria for the diagnosis of RA. The replication samples of European ancestry were derived from a study via electronic health records (EHR).[8] A total of 981 $ACPA^+$ cases and 2,048 controls of European genetic ancestry were genotyped with Sequenom iPLEX at the Broad Institute via methods previously described.[8] Because these RA patients were recruited from the same geographic region as samples from one of our GWASs, we used 129 SNPs that overlapped between the GWAS and the replication study to remove duplicate individuals. Samples in which the proportion of alleles shared with an identity by state of 1 were excluded from the EHR replication dataset, leaving 711 $ACPA^+$ cases and 1,968 controls. The total genotyping rate across the 13 SNPs in these individuals was 97%. The African American sample set consisted of 440 seropositive cases (RF^+ or $ACPA^+$) from the CLEAR (Consortium for the Longitudinal Evaluation of African Americans with Early Rheumatoid Arthritis) registry[10] and 795 controls (kindly provided by Drs. Robert P. Kimberly and Jeffrey C. Edberg) from either the CLEAR Registry or the Birmingham, Alabama area. We genotyped these samples at the Broad Institute by using Sequenom iPLEX. The total genotyping rate across the 13 SNPs in these individuals was 99%. The quality-control (QC) metrics can be found in Table S2. The Japanese dataset consisted of 2,414 cases and 14,245 controls.[12] We generated genotype data with the Illumina HumanHap610-Quad BeadChip, and we performed imputation by using MACH version 1.0.16 and HapMap Phase II JPT+CHB as

Table 2. Characteristics of Samples Included in Our Multiethnic Replication Panel

Study	Geographic Origin	Ethnicity	Case Autoantibody Status	Cases	Controls	Genotyping Platform	Case-Control Stratification Correction
EHR EU	Boston, USA	European Descent	100% ACPA$^+$	711	1,968	Sequenom iPLEX	ancestry informative markers, PC matching, and self-reported ancestry
CLEAR	Southeastern USA	African American	100% RF$^+$ or ACPA$^+$	440	795	Sequenom iPLEX	ethnically and geographically matched
The BioBank Japan Project	Japan	Japanese	76% RF$^+$, 79% ACPA$^+$	2,414	14,245	Illumina HumanHap610-Quad BeadChip	GWAS data PC matching and self-reported ancestry
Korea	Seoul, Korea	Korean	96% ACPA$^+$	801	757	Illumina 550v3/660w	GWAS data PC matching

The following abbreviations are used: GWAS, genome-wide association study; and PC, principal components.

a reference panel (release 24), as previously described.[13] The Korean dataset was genotyped with the Illumina 550v3/660w platform, and 1 Mb regional imputation was performed with BEAGLE and HapMap phase III CHB+JPT as a reference panel.[14] From the Korean dataset, one SNP (rs3890745) could not be imputed because very few SNPs passed QC at this locus in this dataset. Table 3 has a summary of the SNPs used in the multiethnic replication and includes proxy SNPs used in some sample collections.

To test for association, we obtained odds ratios and confidence intervals from unconditional logistic regression in each individual dataset as implemented in SNPTEST v2[15] (for European GWAS sample sets) or PLINK[16] (for all remaining sample sets). We performed meta-analysis by using the inverse-variance fixed method[17] in R version 2.10 in the following three phases: (1) six GWASs and previously genotyped sample sets, all of European descent (Table 1); (2) four multiethnic-replication sample sets (Table 3); and (3) all 18 available datasets (Table 3). Heterogeneity-of-odds ratios across all sample collections were assessed with Cochran's Q method as implemented in R. *Z* scores across each population under study can be found in Table S3.

Of the 13 previously suggestive SNPs investigated (Table 1), seven SNPs replicated at $p < 0.05$ (Table 3). Four of these (rs3890745 at chr1p36, rs11594656 at chr10p15, rs8045689 at chr16p11, and rs2872507 at chr17q12) were significantly below the study-wide Bonferroni corrected p value threshold ($p < 0.0038$) (Table 3). Two SNPs, rs3890745 (at chr1p36) and rs2872507 (at chr17q12), surpassed the genome-wide conservative level of significance in a joint analysis of the multiethnic replication sample sets and previous European GWAS and replication datasets ($p = 2.3 \times 10^{-12}$ and $p = 1.7 \times 10^{-9}$, respectively) consisting of 16,659 cases and 49,174 controls. We note that one SNP, rs8045689, had a lower genotype call rate (90.1%) in the European replication cohort (Table S2) and had evidence of heterogeneity in the multiethnic replication (Table 3).

All SNPs with $p < 0.05$ in replication have been implicated in other immune-mediated diseases, supporting the idea that they represent true positive associations in RA. For the two SNPs that reached genome-wide significance, both are associated with other autoimmune diseases (rs3890745 at chr1p36—Ulcerative Colitis [UC[18] (MIM 605225)] and Celiac disease[19] [MIM 212750]; rs2872507 at chr17q12—Crohn disease[20] [MIM 26600], UC[21] [MIM 26600], type 1 diabetes[22] [T1D (MIM 222100)], asthma[23] [MIM 600807], and primary biliary cirrhosis[24] [MIM 109720]). SNPs at the chr16p11 (rs8045689, combined $p = 7.3 \times 10^{-8}$) and 10p15 (rs11594656, combined $p = 8.46 \times 10^{-6}$) loci have previously been associated with T1D,[25] and the same risk allele predisposes to both RA and T1D. Several SNPs for which results were replicated at $p < 0.05$ in our study are associated with immune-related diseases: rs11203203 at chr21q22 (celiac disease),[3] rs3184504 at chr12q24 (celiac disease and T1D),[26,27] and rs2793108 at chr10p11 (T1D) (May 2009 release of the online T1D database[22]).

We used available GWAS data to fine map the two loci that reached genome-wide significance (Figures 1 and 2). At the chr1p36 locus, the best SNP, rs3890745, is strongly associated with RA risk in both European and Japanese datasets (Figures 1A–1C). This SNP was not genotyped or imputed into our Korean dataset because of the low-density of genotyped SNPs in the region. In both GWAS datasets, this SNP (or SNPs in high linkage disequilibrium [LD]) represents the strongest signal of association. After conditional SNPTEST analysis, no additional signal remains (Figure S1A). Thus, we conclude that the causal variant is in strong LD with rs3890745.

In an attempt to identify the most likely associated gene and any potential causal variants in LD with rs3890745, we performed a series of bioinformatic analyses. First, we used GRAIL[28] to search the region for genes that were most closely related to other established RA risk loci. Using the GRAIL default parameters (CEU [Utah residents with ancestry from northern and western Europe from the CEPH collection] HapMap release 21; PubMed text [Dec. 2006), gene size correction "off"), we used a set of 36 validated RA risk-associated SNPs as "seed regions" (selected from Stahl et al.[2]), and we used our 13 suggestive SNPs as "query regions." The following six genes are in the region of LD: *PANK4* (MIM 606162), *MMEL1* (MIM 18030),

Table 3. Independent Replication from Populations of Diverse Ancestry Including Europeans, African Americans, and East Asians

			Europeans				African Americans				Japanese				Koreans				Multiethnic Replication			Joint Meta-Analysis		
SNP	**Chr.**	**Locus**	**Case**	**Control**	**OR**	**p***	**Case**	**Control**	**OR**	**p***	**Case**	**Control**	**OR**	**p***	**Case**	**Control**	**OR**	**p***	**OR**	**p***	**Q**	**OR**	p_{Joint}	**Q**
rs3890745	1	TNFRSF14	0.71	0.67	1.22	0.002	0.49	0.46	1.14	0.07	0.53	0.50	1.13	5.34×10^{-5}	not imputable				1.14	4.0×10^{-7}	0.60	1.13	$\mathbf{2.3 \times 10^{-12}}$	0.37
rs7543174	1	IL6R	0.20	0.20	0.98	0.62	0.44	0.39	1.21	0.01	0.11	0.12	0.91	0.97	0.16	0.16	1.05	0.34	0.99	0.63	0.04	1.07	3.5×10^{-4}	0.01
rs12746613	1	FCGR2A	0.15	0.14	1.03	0.37	not available				monomorphic				not imputable				1.03	0.37	-	1.11	4.6×10^{-5}	0.21
rs10919563	1	PTPRC	0.87	0.86	1.15	0.06	0.61	0.62	0.98	0.58	0.76	0.77	0.97	0.79	0.76	0.76	1.01	0.47	0.99	0.58	0.38	1.07	1.1×10^{-3}	0.24
rs13119723	4	IL2, IL21	0.15	0.14	1.13	0.92	0.02	0.02	0.90	0.37	monomorphic				not imputable				1.11	0.90	0.48	0.90	9.0×10^{-6}	0.14
rs11594656	10	IL2RA	0.76	0.73	1.19	0.01	0.96	0.96	1.08	0.37	0.96	0.96	1.16	0.04	low information score				1.17	0.001	0.89	1.09	8.5×10^{-6}	0.88
rs2793108	10	ZEB1	0.59	0.60	0.98	0.62	0.38	0.40	0.92	0.83	0.58	0.55	1.11	0.001	0.54	0.56	0.91	0.87	1.05	0.03	0.03	1.07	9.4×10^{-6}	0.39
rs540386	11	TRAF6	0.14	0.14	0.98	0.41	0.74	0.75	0.99	0.46	0.97	0.97	1.07	0.29	0.96	0.95	1.10	0.31	1.01	0.42	0.90	1.10	6.5×10^{-5}	0.15
rs3184504	12	SH2B3	0.52	0.48	1.18	0.004	0.08	0.08	0.99	0.53	monomorphic				not imputable				1.15	0.008	0.31	1.09	3.9×10^{-7}	0.07
rs8045689	16	CD19, NFATC2IP	0.28	0.22	1.40	2.57×10^{-6}	0.05	0.04	1.26	0.12	0.08	0.08	1.03	0.28	0.11	0.09	1.22	0.08	1.16	1.4×10^{-4}	0.01	1.10	7.3×10^{-8}	0.05
rs2872507	17	IKZF3	0.56	0.57	0.96	0.25	0.71	0.76	0.79	0.006	0.72	0.74	0.91	0.003	0.72	0.76	0.82	0.02	0.90	5.3×10^{-5}	0.26	0.91	$\mathbf{1.7 \times 10^{-9}}$	0.74
rs11203203	21	UBASH3A	0.39	0.36	1.12	0.04	0.17	0.16	1.14	0.12	0.04	0.04	0.99	0.57	0.04	0.03	1.35	0.09	1.09	0.03	0.44	1.09	7.7×10^{-7}	0.59
rs5754217	22	UBE2L3	0.20	0.19	1.04	0.31	0.05	0.06	0.85	0.80	0.48	0.47	1.03	0.15	0.45	0.43	1.07	0.22	1.03	0.11	0.75	1.07	1.2×10^{-4}	0.94

We list results for 13 SNPs that were previously suggestive in two GWAS meta-analysis studies but that had not yet reached the conservative genome-wide significance level at $p < 5 \times 10^{-8}$. For each of the sample collections, we report the case-control allele frequencies of allele A1 (as denoted in Table 2), the odds ratio (OR) with respect to A1, and the one-tailed p value (p*) with respect to the directionality in the previous meta-analysis results in Table 1. We then perform a multi-ethnic replication for the independent replication samples as listed in this table, as well as a joint analysis of all independent collections available (*Z* scores for each sample collection are available in Table S3). Underlined p values denote replication below the study-wide Bonferroni corrected p value threshold (<0.0038). p values in bold denote replication below the genome-wide significance threshold ($p < 5 \times 10^{-8}$). Two out of the 13 SNPs genotyped in the European ancestry replication dataset were proxies of the GWAS SNPs (rs793096 at *ZEB1* [LD with rs2793108 $r^2 = 0.82$, $D' = 1$] and rs9621715 at *UBE2L3* [LD with rs5754217 $r^2 = 1$, $D' = 1$]). For the African American dataset, rs9621715 at UBE2L3 and rs793096 at ZEB1 were genotyped. For the Korean dataset, the proxies used were rs1046864 at TRAF6 and rs1932435 at PTPRC. Three SNPs (rs12746613 at FCGR2A, rs13119723 at IL2-21, and rs3184504 at SH2B3) could not be imputed because these SNPs were monomorphic in the reference panel, and one SNP (rs11594656 at IL2RA) had an information score of 0.004 and was excluded from further analysis in this dataset. The following abbreviations are used: Chr., chromosomal region; and Q, Cochran's Q test for heterogeneity of odds ratios.

Please cite this article in press as: Kurreeman et al., Use of a Multiethnic Approach to Identify Rheumatoid- Arthritis-Susceptibility Loci, 1p36 and 17q12, The American Journal of Human Genetics (2012), doi:10.1016/j.ajhg.2012.01.010

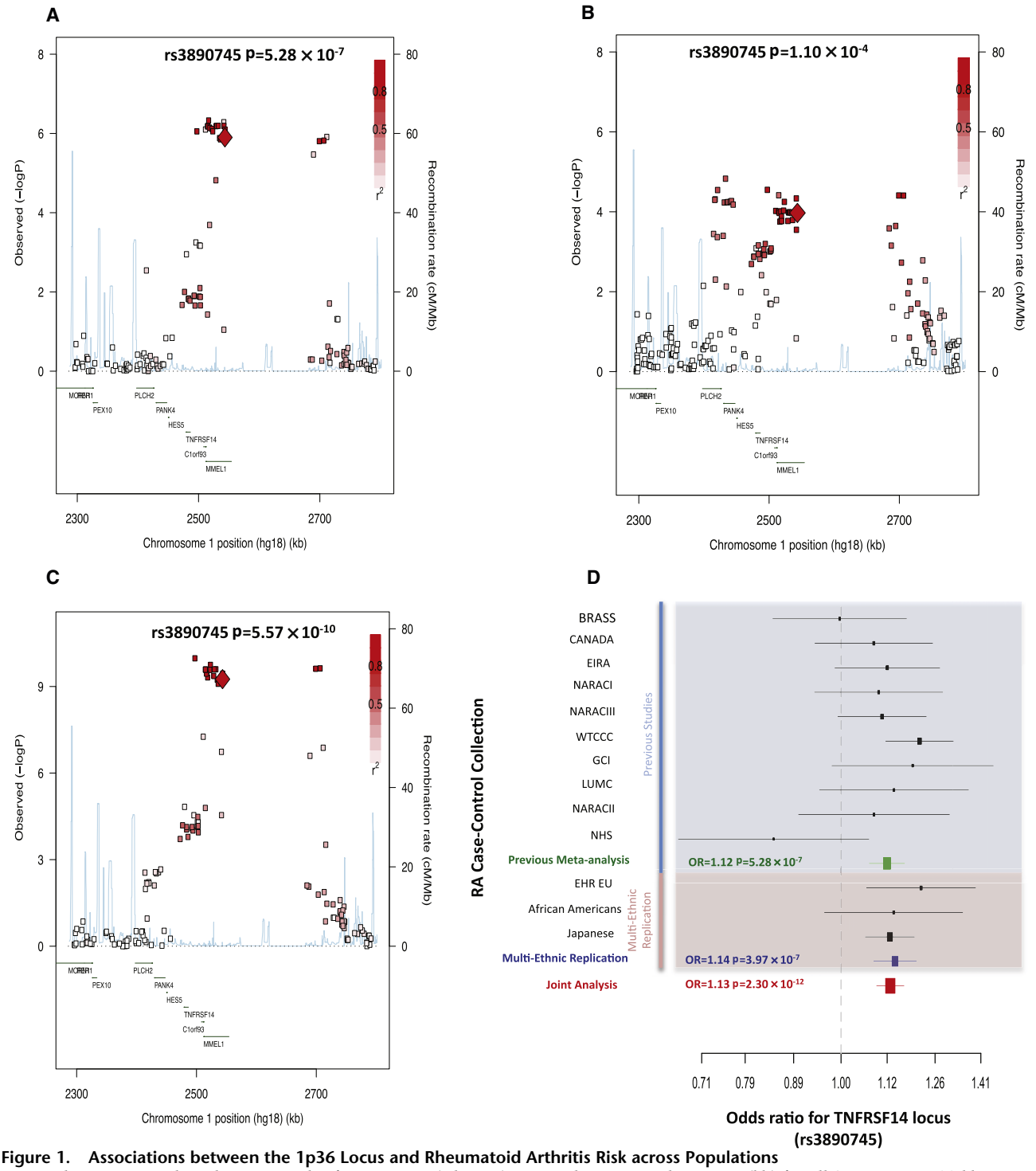

Figure 1. Associations between the 1p36 Locus and Rheumatoid Arthritis Risk across Populations

Regional association plots show strength of association ($-\log_{10}$p) versus chromosomal position (kb) for all SNPs across 500 kb regions centered on the newly validated SNPs (labeled). p values are plotted with diamonds for all SNPs and are shaded white to red by the degree of LD (r^2; see inset) with the SNP (larger red diamond) under investigation in the current study. Local recombination rates estimated from HapMap CEU (cM/Mb, blue line) are plotted against the secondary *y* axis and show recombination hotspots across the region. Labeled green arrows below the plots indicate genes and their orientations.

(A) Associations from the European population.

(B) Associations from the Japanese population.

(C) Meta-analysis of European and Japanese populations.

(D) Forest plot showing association across all available studies. The point estimate of the odds ratio (OR) and 95% confidence intervals (CIs) are shown for each individual study included in the meta-analysis, as well as for a combined analysis (green) across the six GWASs and the previously available replication sample sets, the multiethnic sample sets (blue) novel to this study, and the meta-analysis (red) across all sample sets.

Please cite this article in press as: Kurreeman et al., Use of a Multiethnic Approach to Identify Rheumatoid- Arthritis-Susceptibility Loci, 1p36 and 17q12, The American Journal of Human Genetics (2012), doi:10.1016/j.ajhg.2012.01.010

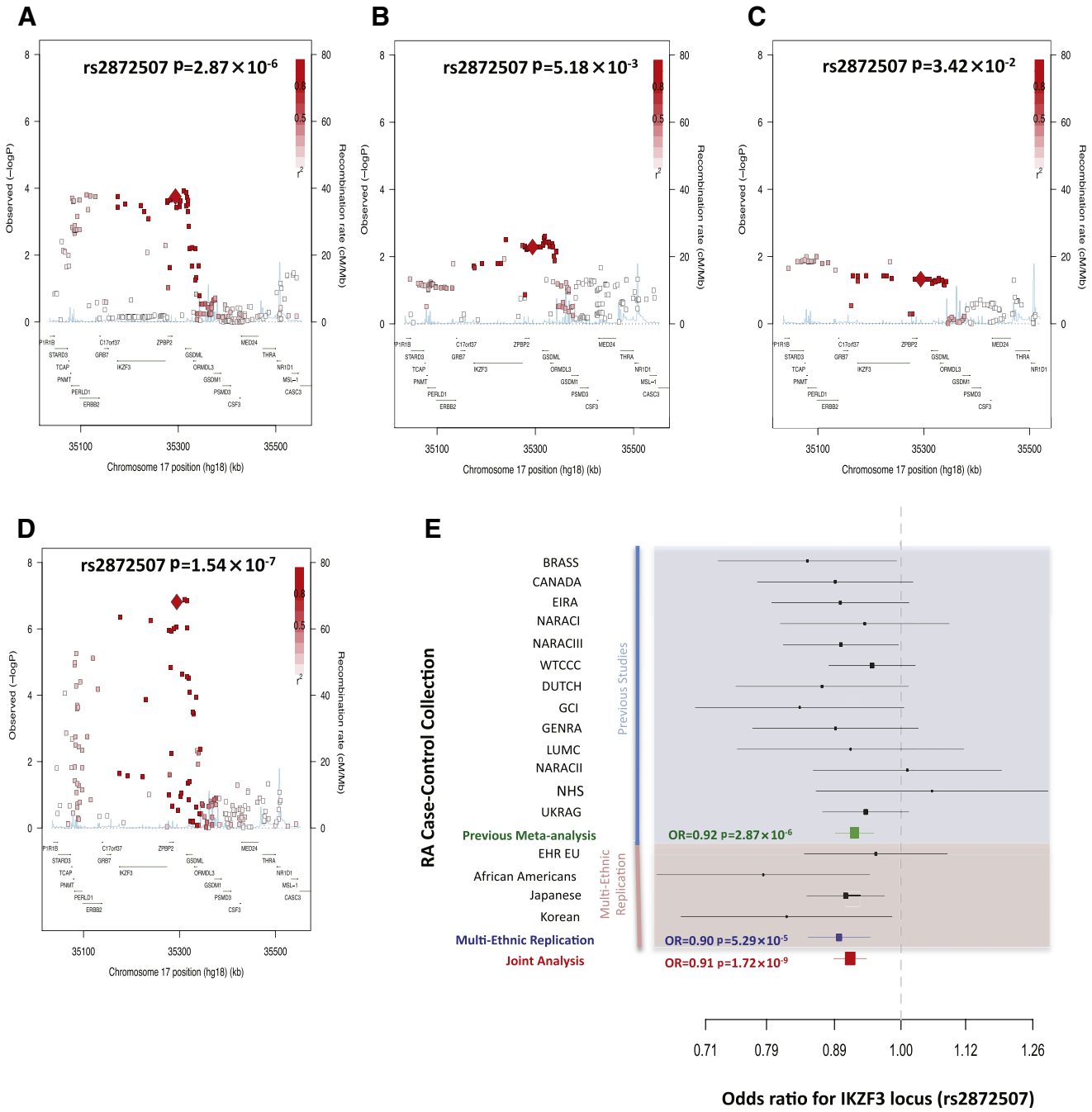

Figure 2. Associations between the 17q12 Locus and Rheumatoid Arthritis Risk across Populations

Regional association plots show strength of association ($-\log_{10}$p) versus chromosomal position (kb) for all SNPs across 500 kb regions centered on the newly validated SNPs (labeled). p values are plotted with diamonds for all SNPs and are shaded white to red by the degree of LD (r^2; see inset) with the SNP (larger red diamond) under investigation in the current study. Local recombination rates estimated from HapMap CEU (cM/Mb, blue line) are plotted against the secondary *y* axis and show recombination hotspots across the region. Labeled green arrows below the plots indicate genes and their orientations.

(A) Associations from the European population.

(B) Associations from the Japanese population.

(C) Associations from the Korean population.

(D) Meta-analysis of European, Japanese, and Korean populations.

(E) Forest plot showing association across all available studies. The point estimate of the odds ratio (OR) and 95% confidence intervals (CIs) are shown for each individual study included in the meta-analysis as well as for a combined analysis (green) across the six GWASs and the previously available replication sample sets, the multiethnic sample sets (blue) novel to this study, and the meta-analysis (red) across all sample sets.

PLCH2 (MIM 612836), *C1orf93*, *HES5* (MIM 607348), and *TNFRSF14* (MIM 602746). GRAIL picked *TNFRSF14* as the gene most likely to have a causal variant in this region ($p_{GRAIL} = 2 \times 10^{-6}$), and no other gene scored significantly at $p_{GRAIL} < 0.05$. This gene is a member of the TNF (tumor necrosis factor)-receptor superfamily and is known to bind to several TRAF (TNF-receptor-associated factors) family members, which might mediate the signal transduction pathways that activate the immune response. We then searched the 1,000 Genomes Project pilot 1 data[29] and catalogued SNPs in high LD with rs3890745 ($r^2 \geq 0.8$). In exon 15 of *MMEL1*, we found one missense SNP (rs3748816; $r^2 = 0.93$, $D' = 1$; methionine to threonine) that was in LD with rs3890745. PolyPhen2[30] predicted that this SNP amino acid change would have benign consequences on the MMEL1 protein. Finally, we used a publicly available genome browser to search for *cis*-acting expression quantitative trait loci (eQTL) on genes in the region. This SNP is a strong eQTL for *MMEL1* (1.03×10^{-20}) in a large dataset of peripheral-blood mononuclear cells (PBMCs).[19] *MMEL1* encodes a member of the neutral endopeptidase (NEP) or membrane metallo-endopeptidase (MME) family. Family members play important roles in pain perception, arterial pressure regulation, phosphate metabolism, and homeostasis. This protein is a type II transmembrane protein and is thought to be expressed as a secreted protein. Determining which gene (or genes) and variants are causal will require functional studies.

We also used GWAS data to fine map the chr17q12 locus marked by rs2872507. Again, the best SNP from our original GWAS on Europeans represented the best signal of association in the GWAS from Japanese and Korean individuals (Figures 2A–2D). In all three GWASs, the strongest signal of association was with rs2872507. After conditional analysis, no additional signal remains (Figure S1B). Thus, we conclude that the causal variant is in strong LD with rs2872507.

We performed the following three similar bioinformatic analyses to identify the most likely causal variant and gene on which it is located at the rs2872507 locus at chr17q12. (1) This region contains 17 genes (Figure 2), of which *IKZF3* [MIM 606221] is the best biological candidate gene identified by GRAIL ($p_{GRAIL} = 2 \times 10^{-5}$), and no other gene scored significantly at $p_{GRAIL} < 0.05$. *IKZF3* (IKAROS family zinc finger 3, also known as Aiolos) has an important function in the regulation and proliferation of B cells.[31] Mice lacking *IKZF3* develop symptoms of human systemic lupus erythematosus (SLE), indicating that normal *IKZF3* function might be necessary for maintaining immune homeostasis and suppressing the development of systemic autoimmune disease.[32] (2) There were three missense SNPs in LD with rs2872507, two of which are in *GSDMB* (Gasdermin B [MIM 611221]). One (rs2305479) of these two is predicted by PolyPhen2 to be probably damaging as a result of an amino acid change from Glycine to Arginine, and the other (rs2305480) is predicted to be benign. *GSDMB* encodes a member of the gasdermin-domain-containing protein family and is highly expressed in the thymus, lymph nodes, and $CD4^+$ and $CD8^+$ T cells. A third missense SNP is rs11557467 and is located in exon 4 of *ZPBP2* (zona pellucida binding protein 2), which is not a strong biological candidate gene for RA. (3) This SNP is a strong eQTL for *ORMDL3* [MIM 610075] and possibly other genes in the region.[33] A recent paper investigated the potential functional consequences of the SNPs in the LD block and identified a proxy for our top hit (rs12936231, r^2 0.91, $D' = 1$) as disrupting CTCF binding and nucleosome occupancy.[33] As with the chr1p36 region, further functional studies are required for identifying the causal variant in the region.

Our results are consistent with similar genetic architecture across the ethnic groups (Figure S2). In particular, we provide evidence of shared risk alleles among Japanese and European individuals, given that these represent the ethnic groups with the largest number of RA cases and controls in our study (Tables 1 and 2). For each of the multiethnic replication sample sets, we used Fisher's method to test whether there was a uniform distribution of the p value across the 13 SNPs genotyped. In all datasets, we observed significantly higher association in individuals of European ancestry ($p_{EHR\_EU} = 1.25 \times 10^{-07}$; $p_{AA} = 0.01$, $p_{JAPAN} = 3.45E^{-06}$; $p_{KOREA} = 0.04$). Within each of the datasets, we observed that six SNPs in the EHR-EU dataset, two SNPs in the AA dataset, four SNPs in the Japanese dataset, and one SNP in the Korean dataset were significant ($p < 0.05$ [corresponding to $Z > 1.65$]), whereas no more than 1 might be expected by chance alone (Figure S3). A summary of the power estimates for each of the sample sets is presented in Figure S4.

We also highlight apparent differences across ethnic groups. First, there are three SNPs (rs12746613 at *FCGR2A*, rs13119723 at *IL2-IL21*, and rs3184504 at *SH2B3*) that are monomorphic among individuals of Asian ancestry but that are polymorphic among individuals of European ancestry. This limits our ability to detect a true positive association in a multiethnic study design and also explains why we were not able to impute this SNP in the Korean GWAS. One SNP in particular is rs3184504 (on chr12 near *SH2B3*), which replicates with $p = 0.004$ among individuals of European ancestry. This same SNP was recently found to be associated with celiac and RA[3]. There is also evidence of heterogeneity in the association at loci that failed to reach combined $p < 5 \times 10^{-8}$ (e.g., rs2793108 and rs7543174). It is possible that heterogeneity is explained by clinical variability across ethnic groups, different patterns of LD between the genotyped marker SNP (Figures S5 and S6) and the underlying causal variant among ethnic populations, or the existence of different causal variants in individuals of different ethnic backgrounds. In these instances, a multiethnic study design does not result in a gain in power. It is also possible that these do not represent true positive associations.

A limitation of our study is highlighted by our efforts to find the causal variant and the gene on which it is located

at the two loci (chr17q12 and 1p36) that reached genome-wide significance. We used GWAS data and 1,000 Genomes Project data to identify a set of equivalent SNPs, but we were not able to pinpoint the causal variant. Similarly, our bioinformatic analyses implicated more than one gene per locus as the gene most likely influenced by the causal variant. Resolving both issues will require detailed functional studies.

Our study has implications beyond the identification of two RA risk loci. It is increasingly recognized that common alleles of small effect can explain a substantial proportion of the hidden heritability of complex traits,[34,35] including the risk of developing RA (Stahl et al., in press). Obtaining sufficient power for identifying these risk alleles will require very large sample sizes. Our study demonstrates that combining GWASs across multiple ethnic groups represents an effective strategy for discovering RA risk loci.

Supplemental Data

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

Acknowledgments

F.K. was supported by the European Community's FP7 Marie Curie grant (Grant Agreement number PIOF-GA-2009-237280). R.M.P. was supported by grants from the National Institutes of Health (R01-AR057108, R01-AR056768, and U01-GM092691) and holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund. S.Y.B., H.S.L., and S.C.B. were supported by the Korea Healthcare technology R&D Project, Ministry for Health and Welfare, Republic of Korea (A111218-11-GM01, A102065).

CLEAR Investigators include: Doyt L. Conn (Emory University), Beth L. Jonas, Leigh F. Callahan (University of North Carolina), Edwin A. Smith (Medical University of South Carolina), Richard D. Brasington, Jr. (Washington University), and Larry W. Moreland (University of Pittsburgh). We would like to thank Robert Kimberly and Jeff Edberg for contributing DNA samples from African American controls.

Received: October 7, 2011
Revised: December 8, 2011
Accepted: January 12, 2012
Published online: February 23, 2012

References

1. MacGregor, A.J., Snieder, H., Rigby, A.S., Koskenvuo, M., Kaprio, J., Aho, K., and Silman, A.J. (2000). Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum. *43*, 30–37.
2. Stahl, E.A., Raychaudhuri, S., Remmers, E.F., Xie, G., Eyre, S., Thomson, B.P., Li, Y., Kurreeman, F.A., Zhernakova, A., Hinks, A., et al; BIRAC Consortium; YEAR Consortium. (2010). Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat. Genet. *42*, 508–514.
3. Zhernakova, A., Stahl, E.A., Trynka, G., Raychaudhuri, S., Festen, E.A., Franke, L., Westra, H.J., Fehrmann, R.S., Kurreeman, F.A., Thomson, B., et al. (2011). Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet. *7*, e1002004.
4. Chen, R., Stahl, E.A., Kurreeman, F.A., Gregersen, P.K., Siminovitch, K.A., Worthington, J., Padyukov, L., Raychaudhuri, S., and Plenge, R.M. (2011). Fine mapping the TAGAP risk locus in rheumatoid arthritis. Genes Immun. *12*, 314–318.
5. Raychaudhuri, S., Remmers, E.F., Lee, A.T., Hackett, R., Guiducci, C., Burtt, N.P., Gianniny, L., Korman, B.D., Padyukov, L., Kurreeman, F.A., et al. (2008). Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat. Genet. *40*, 1216–1223.
6. Raychaudhuri, S., Thomson, B.P., Remmers, E.F., Eyre, S., Hinks, A., Guiducci, C., Catanese, J.J., Xie, G., Stahl, E.A., Chen, R., et al; BIRAC Consortium; YEAR Consortium. (2009). Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat. Genet. *41*, 1313–1318.
7. Kochi, Y., Okada, Y., Suzuki, A., Ikari, K., Terao, C., Takahashi, A., Yamazaki, K., Hosono, N., Myouzen, K., Tsunoda, T., et al. (2010). A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. Nat. Genet. *42*, 515–519.
8. Kurreeman, F., Liao, K., Chibnik, L., Hickey, B., Stahl, E., Gainer, V., Li, G., Bry, L., Mahan, S., Ardlie, K., et al. (2011). Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. Am. J. Hum. Genet. *88*, 57–69.
9. Freudenberg, J., Lee, H.S., Han, B.G., Shin, H.D., Kang, Y.M., Sung, Y.K., Shim, S.C., Choi, C.B., Lee, A.T., Gregersen, P.K., and Bae, S.C. (2011). Genome-wide association study of rheumatoid arthritis in Koreans: Population-specific loci as well as overlap with European susceptibility loci. Arthritis Rheum. *63*, 884–893.
10. Hughes, L.B., Reynolds, R.J., Brown, E.E., Kelley, J.M., Thomson, B., Conn, D.L., Jonas, B.L., Westfall, A.O., Padilla, M.A., Callahan, L.F., et al. (2010). Most common single-nucleotide polymorphisms associated with rheumatoid arthritis in persons of European ancestry confer risk of rheumatoid arthritis in African Americans. Arthritis Rheum. *62*, 3547–3553.
11. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H., and Goldstein, D.B. (2010). Rare variants create synthetic genome-wide associations. PLoS Biol. *8*, e1000294.
12. Nakamura, Y. (2007). The BioBank Japan Project. Clin. Adv. Hematol. Oncol. *5*, 696–697.
13. Okada, Y., Hirota, T., Kamatani, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Higasa, K., Yamaguchi-Kabata, Y., Hosono, N., Nalls, M.A., et al. (2011). Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. PLoS Genet. *7*, e1002067.
14. Browning, B.L., and Browning, S.R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. *84*, 210–223.
15. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. *39*, 906–913.
16. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

17. de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S., and Voight, B.F. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. *17* (R2), R122–R128.
18. Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D'Amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A., et al. (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat. Genet. *43*, 246–252.
19. Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A., Adány, R., Aromaa, A., et al. (2010). Multiple common variants for celiac disease influencing immune gene expression. Nat. Genet. *42*, 295–302.
20. Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., Brant, S.R., Silverberg, M.S., Taylor, K.D., Barmada, M.M., et al; NIDDK IBD Genetics Consortium; Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat. Genet. *40*, 955–962.
21. McGovern, D.P., Gardet, A., Törkvist, L., Goyette, P., Essers, J., Taylor, K.D., Neale, B.M., Ong, R.T., Lagacé, C., Li, C., et al; NIDDK IBD Genetics Consortium. (2010). Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nat. Genet. *42*, 332–337.
22. Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C., et al; Type 1 Diabetes Genetics Consortium. (2009). Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat. Genet. *41*, 703–707.
23. Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., von Mutius, E., Farrall, M., Lathrop, M., and Cookson, W.O.; GABRIEL Consortium. (2010). A large-scale, consortium-based genomewide association study of asthma. N. Engl. J. Med. *363*, 1211–1221.
24. Liu, X., Invernizzi, P., Lu, Y., Kosoy, R., Lu, Y., Bianchi, I., Podda, M., Xu, C., Xie, G., Macciardi, F., et al. (2010). Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. Nat. Genet. *42*, 658–660.
25. Lowe, C.E., Cooper, J.D., Brusko, T., Walker, N.M., Smyth, D.J., Bailey, R., Bourget, K., Plagnol, V., Field, S., Atkinson, M., et al. (2007). Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. Nat. Genet. *39*, 1074–1082.
26. Hunt, K.A., Zhernakova, A., Turner, G., Heap, G.A., Franke, L., Bruinenberg, M., Romanos, J., Dinesen, L.C., Ryan, A.W., Panesar, D., et al. (2008). Newly identified genetic risk variants for celiac disease related to the immune response. Nat. Genet. *40*, 395–402.
27. Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S.F., Payne, F., et al; Genetics of Type 1 Diabetes in Finland; Wellcome Trust Case Control Consortium. (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat. Genet. *39*, 857–864.
28. Raychaudhuri, S., Plenge, R.M., Rossin, E.J., Ng, A.C., Purcell, S.M., Sklar, P., Scolnick, E.M., Xavier, R.J., Altshuler, D., and Daly, M.J.; International Schizophrenia Consortium. (2009). Identifying relationships among genomic disease regions: Predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. *5*, e1000534.
29. Cirulli, E.T., and Goldstein, D.B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat. Rev. Genet. *11*, 415–425.
30. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
31. Schmitt, C., Tonnelle, C., Dalloul, A., Chabannon, C., Debré, P., and Rebollo, A. (2002). Aiolos and Ikaros: Regulators of lymphocyte development, homeostasis and lymphoproliferation. Apoptosis *7*, 277–284.
32. Sun, J., Matthias, G., Mihatsch, M.J., Georgopoulos, K., and Matthias, P. (2003). Lack of the transcriptional coactivator OBF-1 prevents the development of systemic lupus erythematosus-like phenotypes in Aiolos mutant mice. J. Immunol. *170*, 1699–1706.
33. Verlaan, D.J., Berlivet, S., Hunninghake, G.M., Madore, A.M., Larivière, M., Moussette, S., Grundberg, E., Kwan, T., Ouimet, M., Ge, B., et al. (2009). Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. Am. J. Hum. Genet. *85*, 377–393.
34. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P.; International Schizophrenia Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752.
35. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. *43*, 519–525.