

ImmunoChip Identifies Novel, and Replicates Known, Genetic Risk Loci for Rheumatoid Arthritis in Black South Africans

Nimmisha Govind,¹ Ananyo Choudhury,² Bridget Hodgkinson,¹ Claudia Ickinger,¹ Jacqueline Frost,³ Annette Lee,⁴ Peter K Gregersen,⁴ Richard J Reynolds,⁵ S Louis Bridges Jr,⁵ Scott Hazelhurst,^{2,6} Michèle Ramsay,^{2,3} and Mohammed Tikly^{1,2}

¹Division of Rheumatology, Faculty of Health Sciences, ²Sydney Brenner Institute for Molecular Bioscience, and ³Division of Human Genetics, National Health Laboratory Service, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; ⁴Feinstein Institute for Medical Research, Manhasset, New York, United States of America; ⁵Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, Alabama, United States of America; ⁶School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa

The aim of this study was to identify genetic variants associated with rheumatoid arthritis (RA) risk in black South Africans. Black South African RA patients ($n = 263$) were compared with healthy controls ($n = 374$). Genotyping was performed using the ImmunoChip, and four-digit high-resolution human leukocyte antigen (HLA) typing was performed by DNA sequencing of exon 2. Standard quality control measures were implemented on the data. The strongest associations were in the intergenic region between the *HLA-DRB1* and *HLA-DQA1* loci. After conditioning on *HLA-DRB1* alleles, the effect in the rest of the extended major histocompatibility (MHC) diminished. Non-HLA single nucleotide polymorphisms (SNPs) in the intergenic regions *LOC389203/RBPJ*, *LOC100131131/IL1R1*, *KIAA1919/REV3L*, *LOC643749/TRAF3IP2*, and SNPs in the intron and untranslated regions (UTR) of *IRF1* and the intronic region of *ICOS* and *KIAA1542* showed association with RA ($p < 5 \times 10^{-5}$). Of the SNPs previously associated with RA in Caucasians, one SNP, rs874040, locating to the intergenic region *LOC389203/RBPJ* was replicated in this study. None of the variants in the *PTPN22* gene was significantly associated. The seropositive subgroups showed similar results to the overall cohort. The effects observed across the HLA region are most likely due to *HLA-DRB1*, and secondary effects in the extended MHC cannot be detected. Seven non-HLA loci are associated with RA in black South Africans. Similar to Caucasians, the intergenic region between *LOC38920* and *RBPJ* is associated with RA in this population. The strong association of the R620W variant of the *PTPN22* gene with RA in Caucasians was not replicated since this variant was monomorphic in our study, but other SNP variants of the *PTPN22* gene were also not associated with RA in black South Africans, suggesting that this locus does not play a major role in RA in this population.

Online address: <http://www.molmed.org>
doi: 10.2119/molmed.2014.00097

INTRODUCTION

Rheumatoid arthritis (RA) is a complex autoimmune disease affecting 0.5% to 1% of the population worldwide. This is characterized by chronic inflammation of synovial joints resulting in progressive joint destruction. The etiology of RA re-

mains elusive; however it is thought to occur in a genetically predisposed individual who is exposed to a set of environmental triggers. The heritability of RA is estimated to be as high as 50% to 60% (1).

In recent years, genome-wide association studies (GWASs) have made great

strides in identifying novel loci associated with RA. Despite these discoveries, less than 50% of the heritability can be explained. GWASs also have uncovered that various autoimmune diseases share similar risk loci (2). Thus, aggregating diseases with similar pathogenesis is a logical and cost effective method to identify associated loci that overlap various autoimmune diseases. The ImmunoChip consortium designed a genotyping array, an Illumina Infinium array with ~196,000 single nucleotide polymorphisms (SNPs), from 186 loci previously associated with 12 autoimmune diseases identified from GWASs. The success of the ImmunoChip is evidenced by novel risk loci, not previously associated with a specific pheno-

Address correspondence to Nimmisha Govind, Division of Rheumatology, Chris Hani Baragwanath Academic Hospital, University of the Witwatersrand, Johannesburg, South Africa. Phone: +27 11-933-9377; Fax: +27 11-933 8937; E-mail: nimmisha.govind@gmail.com. Submitted May 4, 2014; Accepted for publication July 3, 2014; Epub (www.molmed.org) ahead of print July 7, 2014.

The Feinstein Institute
for Medical Research 

Empowering Imagination. Pioneering Discovery.®

type, being identified in numerous autoimmune diseases (3–8) including RA (9).

Genetic differences underlie the risk for RA between the major ethnic groups. The only genetic variants that have consistently been shown to confer a risk for RA across all ethnic groups is the *HLA-DRB1* alleles conferring a third of the genetic risk. The non-*HLA* genes, however, show less consistency across ethnic groups. The R620W variant of the *PTPN22* gene is strongly and consistently associated with RA in Europeans (10,11) but not in Asians or black South Africans (12–14). In contrast, specific haplotypes of the *PADI4* gene confer risk for RA in Asians (15–18) and less or no risk in Europeans (9,18–21). However a haplotype in the *STAT4* gene is associated with RA in both Europeans and Asians (22,23).

Most large-scale genetic studies have been done in Europeans and Asians, and very few RA risk loci have been studied in Africans. The strong association with the *HLA-DRB1* region has been replicated repeatedly (24–27). Ninety-two percent of black South Africans with RA carry at least one copy of the shared epitope alleles (27) which contrasts with only ~30% in West Africans from Cameroon (28), demonstrating significant heterogeneity within Africans.

To date, there has been no large-scale genetic study performed on black Africans with RA. In view of the observed heterogeneity between and within ethnic groups, our aim was to examine the role of known genetic loci previously associated with RA and to identify novel risk loci in black South Africans with RA.

MATERIALS AND METHODS

Consenting patients were recruited from a single center, the Chris Hani Baragwanath Academic Hospital, Soweto, Johannesburg (n = 414). All patients were unrelated and unselected, fulfilled the American College of Rheumatology 1987 criteria for RA (29) and were over the age of 18 years at disease onset. Patients were considered as “black” South Africans if they self reported all four grandparents as being black South Africans. The controls

were geographically and ethnically matched to the cases (n = 407). They were recruited from the staff of the hospital or from the outpatients department. These were patients with minor trauma and no history of inflammatory joint pain or autoimmune diseases. The study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (M10707).

Serology Tests

Rheumatoid factor (RF) (composite IgM, IgG, IgA) was assayed by nephelometry (Siemens Healthcare Diagnostics, BN Prospec Nephelometer, Newark, DE, USA). Anti-CCP (aCCP) was measured using a second-generation immunofluorimetric assay with the Immunocap 250 system and reagents and controls provided by the manufacturer (Phadia AB, Uppsala, Sweden). Rheumatoid factor and aCCP were considered positive when the concentrations were greater than 15 IU/mL and 10 U/mL, respectively.

Genotyping

ImmunoChip. Genotyping using the ImmunoChip was performed at the Feinstein Institute for Medical Research, Manhasset, NY, USA. Genotype clustering was performed using the default Illumina cluster file (ImmunoChip_Gentrain_June2010.egt) and manifest file (ImmunoBeadChip_11419691.bpm) (NCBI build 36) using the GenTrain2 clustering algorithm. Genotype calling was done using the Genotyping Module of the GenomeStudio Data Analysis Software package. Markers with a significance for association of $p \leq 5 \times 10^{-5}$ were considered significant (5) and the cluster plots were manually inspected.

High resolution HLA DRB1 genotyping. Four-digit high resolution HLA typing was performed by DNA sequencing of exon 2, using the AlleleSEQR HLA DRB1 reagent kit and protocol (Atria Genetics, South San Francisco, CA, USA) at the Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL, USA.

After polymerase chain reaction amplification of HLA DRB1 exon 2 from genomic DNA, forward and reverse cycle sequencing was performed, and the resulting fragments were collected and analyzed on an ABI 377 automated sequencer (Applied Biosystems, Foster City, CA, USA). An additional sequence reaction was performed to analyze the GTG (valine) motif of codon 86 sequences, thus enabling resolution of ambiguous results for some exon 2 sequences. The sequences were analyzed using Assign software (Conexio Genomics, Fremantle, Western Australia, Australia), which enables assignment of genotypes based on a library file of *HLA-DRB1* alleles (30). This method detects all of the SE-positive alleles.

Quality control. Sample quality control (QC) was done initially, followed by SNP QC and population structure analysis. Sample quality control included preprocessing of data in which poor performing samples (genotype call rate <90%) were removed. Thereafter, samples were excluded if the genotype missingness was $\geq 6\%$, the recorded sex differed from the genotype inferred sex, or if they were duplicated and related. Cryptic relatedness was assessed by estimating identity by state (IBS) statistic and an IBS > 0.95 PIHAT < 0.05 using PLINK v1.07, and was used to exclude individuals. Only a single pair of related individuals was found with this cutoff and one of these two individuals was randomly excluded from further analysis. Single nucleotide polymorphisms were excluded if the per SNP missing genotype call rate was $\geq 5\%$ in either cases or controls, if they had a strong difference in missingness among cases and control ($P < 10^{-3}$), they were monomorphic or with a minor allele frequency (MAF) < 0.05, the GenCall (GC) score < 0.15, they were sex chromosome markers or duplicated markers or there was extensive deviation from Hardy-Weinberg equilibrium (HWE) ($p < 5 \times 10^{-7}$) in the controls (31).

Population structure was analyzed using Eigenstat (32) structure and principal component (PC) analyses. The ge-

nomical inflation factor in the final association test was found to be 1.14 (based on median χ^2 test). Plots were constructed based on a comparison with HapMap 3 data (<http://hapmap.ncbi.nlm.nih.gov/>). From the PC analysis (using the first five principal components, weighted by Eigenvalue), we computed the center of the study group and computed the average distance of each individual to the center (average 0.0113). Individuals with a distance greater than 0.017 were excluded from the analysis. This cutoff was chosen by plotting the distribution and choosing the inflection point.

Statistical Analysis

The χ^2 test was performed to determine differences in allele frequencies between patients and controls. Differences were considered to be significant as follows: in the case of HLA loci $p < 5 \times 10^{-8}$ was used and, as defined previously, an *a priori* significance threshold of $p < 5 \times 10^{-5}$ was used for novel RA-associated SNPs in this study (5) and $p < 0.05$ was considered significant for replication.

Allele frequencies were compared between RA patients and controls by the odds ratio (OR) with a 95% confidence interval (CI). *P* values reported are from 2×2 contingency table analyses of counts of minor and major alleles with case-control status and were based on the χ^2 test or Fisher exact test. We also used logistic regression to test single marker association in the extended MHC (Chr 6: 26 Mb to 34 Mb) after partialing out the effects of variability explained by *HLA-DRB1*. The model included as variables the number of copies of each *HLA-DRB1* allele except for *HLA-DRB1* 11:01, which was treated as the referent. In addition, the model included the number of minor alleles of the extended MHC SNP marker to be tested for association, conditional on the *HLA-DRB1* alleles. These conditional analyses were performed to assess the independent effect of the risk *HLA-DRB1* alleles in the extended MHC.

All supplementary materials are available online at www.molmed.org.

RESULTS

After quality control, 263 patients and 374 controls were tested for association using 103,770 SNPs.

The majority of the cases were females 235/263 (89%) with a mean disease duration of 10.6 (SD = 7.3) years. Among those tested, a high proportion were RF (240/254) and aCCP (186/207) positive.

The ancestry informative markers and principle component analyses (PCA) showed a distinction between the black South African RA cases and controls from Caucasians (CEU), West Africans that is, Yoruba of Nigeria (YRI) and the East Africans, the Luhya (LWK) and Maasai (MKK) tribes of Kenya (Supplementary Figure 2). The majority of the samples formed a homogenous cluster; however, some of the cases and controls showed admixture with two other populations encountered in South Africa, namely representative populations for Caucasians and Gujarati Indians, and were excluded from the analysis (33). Thirty seven samples (5 controls and 32 cases) were therefore excluded in the associations analyses based on these findings.

HLA Associations

In total, 77 SNPs reached genome-wide significance ($p < 5 \times 10^{-8}$) in this study (Figure 1), most of which were in the HLA region on chromosome 6. The strongest associations were with one SNP in the intronic region of the *HLA-DRB5* (rs34083746, OR = 6.15, $p = 1.31 \times 10^{-25}$) gene and three SNPs in the intergenic region *HLA DRB1* | *HLA DQA1* (rs3104413, OR = 3.88, $p = 5.49 \times 10^{-21}$; rs3129769, OR = 3.91, $p = 4.60 \times 10^{-21}$; rs6931277, OR = 3.97, $p = 1.03 \times 10^{-21}$). Of the significantly associated SNPs on chromosome 6, 60 SNPs locate to the *HLA DR* or *DQ* regions or the intergenic region between these two genes and 10 SNPs located to genes outside the HLA class II region on chromosome 6 (HLA-associated genes) (Supplementary Table 1).

Four *HLA DRB1* alleles were associated with a risk for RA in black South Africans (*0401, OR = 4.0 [2.5–6.5], $p < 0.0001$;

*0404, OR = 6.9 [3.9–12.2], $p < 0.0001$; *0405, OR = 4.96 [1.6–15.2], $p = 0.0018$; *1001, OR = 1.8 [1.0–3.3], $p = 0.039$). Three alleles conferred protection for RA (*1101, OR = 0.5 [0.4–0.8], $p = 0.0008$); *1301, OR = 0.6 [0.4–0.8], $p = 0.004$; *1302, OR = 0.6 [0.4–1.0], $p = 0.06$) (Table 1). The correlation coefficient for the effect sizes of the significantly associated alleles between Europeans and black South Africans was 0.61 (Supplementary Figure 3).

Single nucleotide polymorphisms in the HLA-associated genes that reached genome-wide significance locate to the intergenic regions *LOC442175* | *ZNF165* (rs149974, OR = 2.2, $p = 1.64 \times 10^{-8}$), *BTNL2* | *HLA-DRA* (rs6932542, OR = 0.5, $p = 2.3 \times 10^{-9}$; rs5007263, OR = 0.50, $p = 2.8 \times 10^{-9}$; rs5007259, OR = 0.50, $p = 2.8 \times 10^{-9}$; rs9268507, OR = 0.51, $p = 3.67 \times 10^{-9}$; rs5007265, OR = 0.51, $p = 4.55 \times 10^{-9}$; rs4502931, OR = 0.52, $p = 1.67 \times 10^{-8}$; rs6926737, OR = 0.52, $p = 4.28 \times 10^{-9}$), the coding region of *CCHCR1* (rs130071, OR = 1.9, $p = 8.69 \times 10^{-8}$) and the intergenic region *PSMB9* | *HLA-DMB* (rs241406, OR = 7.8, $p = 7.89 \times 10^{-10}$).

However, after conditioning on the *HLA-DRB1* alleles, the effects of all significantly associated SNPs in the extended MHC diminished (Figure 2).

The study is underpowered to identify multiple independent effects without any prior hypothesis of association. The effects observed over the HLA region are most likely due to *HLA-DRB1*, and secondary effects cannot be detected.

Non-HLA Associations

A total of 19 non-HLA SNPs locating to seven loci reached a statistical significance of $p < 5 \times 10^{-5}$ (see Figure 1, Table 2), 1 SNP (rs36110812, OR = 1.60) in the intergenic region *LOC38920* | *RBPJ* on chromosome 4 and four SNPs (rs12470623, OR = 0.61; rs6752379, OR = 0.61, rs13001315, OR = 0.62; rs11123911, OR = 0.62) in the intergenic region *LOC100131131* | *IL1R1* on chromosome 2. A further five SNPs in the intronic region and three SNPs in the UTR of *IRF1* and SNPs in the intronic region of *ICOS* (rs6761201, OR = 0.47) and *KIAA1542*

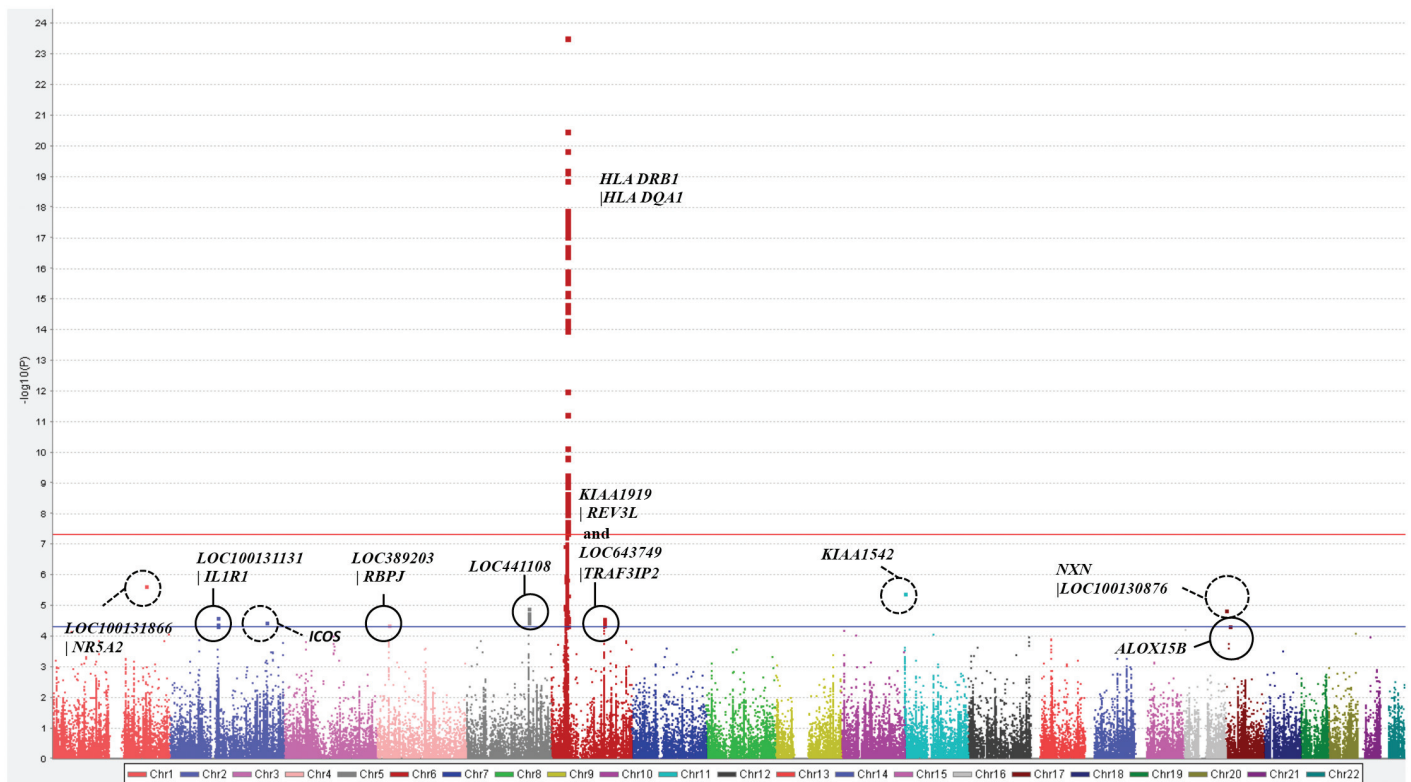


Figure 1. Manhattan plot to highlight the Immuchip loci associated with RA in a black South African study. The strongest associations locate to chromosome 6, the HLA class II loci. Seven non-HLA loci reached statistical significance ($p < 5 \times 10^{-5}$).

(rs12421158, OR = 1.72) reached significance at this level. On chromosome 6, three SNPs locating to the intergenic *KIAA1919|REV3L* and one SNP between *LOC643749* and *TRAF3IP2* showed significance independent of the HLA region. At the majority of regions, we saw a tight cluster of highly correlated vari-

ants (Figure 3). The LocusZoom plots might not represent accurate LD patterns in black South Africans, since data on this population are not available in the 1000 Genomes Project.

Genomic regions and genes within the region are shown in the lower panel (Figures 3A–E). The blue lines show recombination rates within each of the regions. The filled shapes (circles, rectangles and so on) represent the *P* value for SNPs in the region. The shapes signify the function of the SNP based on its localization with respect to nearby genes. Different shapes and their functional implications are summarized in Figure 3F. The purple SNP represents the SNP, which is searched for (shown at the top of the plot) other SNPs in the region are colored depending on their degree of correlation (r^2) with the searched SNP. The degrees of correlation were estimated using LocusZoom on the basis of African population data in the 1000 Genomes Project.

One SNP, rs874040, locating to the intergenic region *LOC389203|RBPJ* was previously associated with RA in Caucasians. In black South Africans, this SNP reached the statistically significant level for replication ($p = 0.001248$, OR = 1.45).

Three other SNPs reached significance in the intergenic regions *LOC100131866|NR5A2* and *NXN|LOC100130876* and in the coding region of *ALOX15B*; however, these were isolated SNPs (Supplementary Figure 1) and were therefore not considered further.

At a significance of $p < 5 \times 10^{-4}$, a further seven new loci were identified (Supplementary Table 2), the intergenic regions *CTLA4|ICOS*, *TNFAIP3|PERP*, *RSPH3|TAGAP*, *IL18RAP|SLC9A4*, and *IL1R2|LOC100131131* and the intronic region of *IL23R* and *IL1R1*.

This study had more than 80% power for allele frequencies ≥ 0.05 to detect effect

Table 1. The OR and 95% CI of the classical HLA DRB1 alleles in black South Africans and Europeans.

<i>HLA-DRB1</i> allele	Black South Africans, OR (95% CI)	Europeans, OR (95% CI)
*0401	4.0 (2.5–6.5)	4.14 (3.86–4.44)
*0404	6.9 (3.9–12.2)	3.17 (2.83–3.54)
*0405	4.96 (1.6–15.2)	2.31 (1.77–3.01)
*1001	1.8 (1.0–3.3)	2.53 (2.04–3.14)
*1101	0.5 (0.4–0.8)	0.44 (0.38–0.52)
*1301	0.6 (0.4–0.8)	0.28 (0.24–0.33)
*1302	0.6 (0.4–1.0)	0.29 (0.23–0.38)

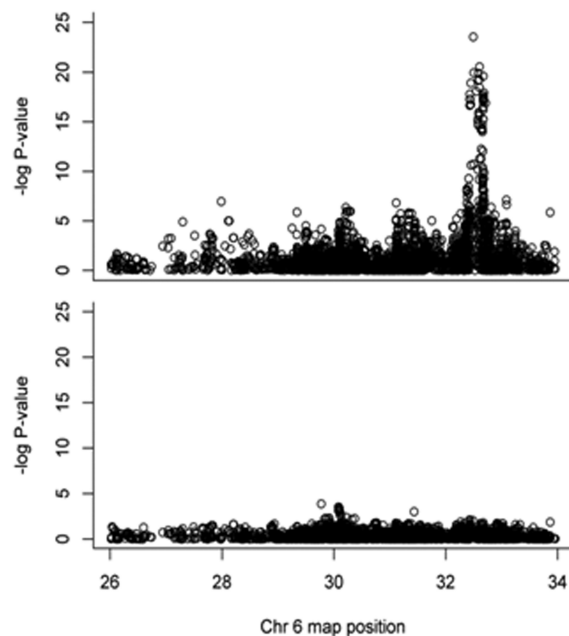


Figure 2. Conditional analyses on the *HLA-DRB1* alleles result in diminished effect in the extended MHC.

sizes of 1.9 and higher. Interestingly, many of the previously associated SNPs were found to be either monomorphic or of lower frequency in our study group.

However, despite the small sample size, we were adequately powered to detect significance for similar effect sizes for several SNPs in the *PTPN22* gene. None

of the SNPs with reasonable frequency was significantly associated with RA. Although the most highly associated allele in European populations is monomorphic (rs2476601 [R620W]) in this study, other SNPs in and around the *PTPN22* gene did not show association, thereby excluding *PTPN22* from making a notable contribution to RA susceptibility in this African cohort.

The strength of association of the significantly associated HLA and non-HLA SNPs showed similar results in the overall cohort and the seropositive (aCCP and RF) subgroups (results not shown).

DISCUSSION

This study is the first large scale genetic project performed on a non-admixed African population with RA. It confirms that the strongest genetic association lies in the HLA class II region of chromosome 6. In addition, several non-HLA associations were observed, including SNPs in the intergenic regions *LOC389203* | *RBPJ*, *LOC100131131* | *IL1R1*, *KIAA1919* | *REV3L*, and *LOC643749* | *TRAF3IP2*, and SNPs in the intron and UTR of *IRF1* and the in-

Table 2. Non-HLA SNPs that reached a statistical significance of $p < 5 \times 10^{-5}$.

CHR	SNP	A1	MAF in cases	MAF in controls	P	OR (95% CI)	Gene	Region
1	rs1325190	A	0.08	0.17	2.30E-06	0.42	<i>LOC100131866</i> <i>NR5A2</i>	Intergenic
2	rs12470623	A	0.34	0.46	2.51E-05	0.61(0.48–0.77)	<i>LOC100131131</i> <i>IL1R1</i>	Intergenic
2	rs6752379	A	0.34	0.46	2.51E-05	0.61(0.48–0.77)	<i>LOC100131131</i> <i>IL1R1</i>	Intergenic
2	rs13001315	A	0.38	0.50	4.14E-05	0.62 (0.50–0.78)	<i>LOC100131131</i> <i>IL1R1</i>	Intergenic
2	rs11123911	A	0.38	0.49	4.61E-05	0.62 (0.50–0.78)	<i>LOC100131131</i> <i>IL1R1</i>	Intergenic
2	rs6761201	G	0.08	0.16	3.45E-05	0.47 (0.33–0.68)	<i>ICOS</i>	Intron
4	rs36110812	G	0.48	0.37	4.25E-05	1.60 (1.28–2.01)	<i>LOC389203</i> <i>RBPJ</i>	Intergenic
5	rs2405528	A	0.17	0.27	2.16E-05	0.55 (0.41–0.72)	<i>IRF1</i>	Intron
5	rs886286	G	0.17	0.27	1.72E-05	0.54 (0.41–0.72)	<i>IRF1</i>	Intron
5	rs757105	A	0.17	0.27	1.93E-05	0.55 (0.41–0.72)	<i>IRF1</i>	Intron
5	rs2522047	A	0.16	0.27	1.26E-05	0.53 (0.40–0.71)	<i>IRF1</i>	Intron
5	rs2706393	A	0.17	0.27	1.93E-05	0.55 (0.41–0.72)	<i>IRF1</i>	Intron
5	rs2522050	A	0.17	0.27	2.71E-05	0.55 (0.42–0.73)	<i>IRF1</i>	UTR
5	rs2706395	T	0.17	0.27	1.93E-05	0.55 (0.40–0.72)	<i>IRF1</i>	UTR
5	rs2522056	A	0.15	0.24	3.42E-05	0.54 (0.40–0.72)	<i>IRF1</i>	UTR
6	rs173286	G	0.47	0.36	3.27E-05	1.62 (1.29–2.03)	<i>KIAA1919</i> <i>REV3L</i>	Intergenic
6	rs25638	I	0.47	0.36	3.27E-05	1.62 (1.29–2.03)	<i>KIAA1919</i> <i>REV3L</i>	Intergenic
6	rs240959	A	0.44	0.33	2.73E-05	1.64 (1.30–2.06)	<i>KIAA1919</i> <i>REV3L</i>	Intergenic
6	rs71562296	A	0.12	0.20	4.36E-05	0.51 (0.37–0.71)	<i>LOC643749</i> <i>TRAF3IP2</i>	Intergenic
11	rs12421158	A	0.47	0.34	3.94E-06	1.72 (1.36–2.16)	<i>KIAA1542</i>	Intron
17	rs7218037	G	0.46	0.34	1.45E-05	1.66	<i>NXN</i> <i>LOC100130876</i>	Intergenic
17	rs9904554	A	0.37	0.27	4.78E-05	1.64	<i>ALOX15B</i>	Coding

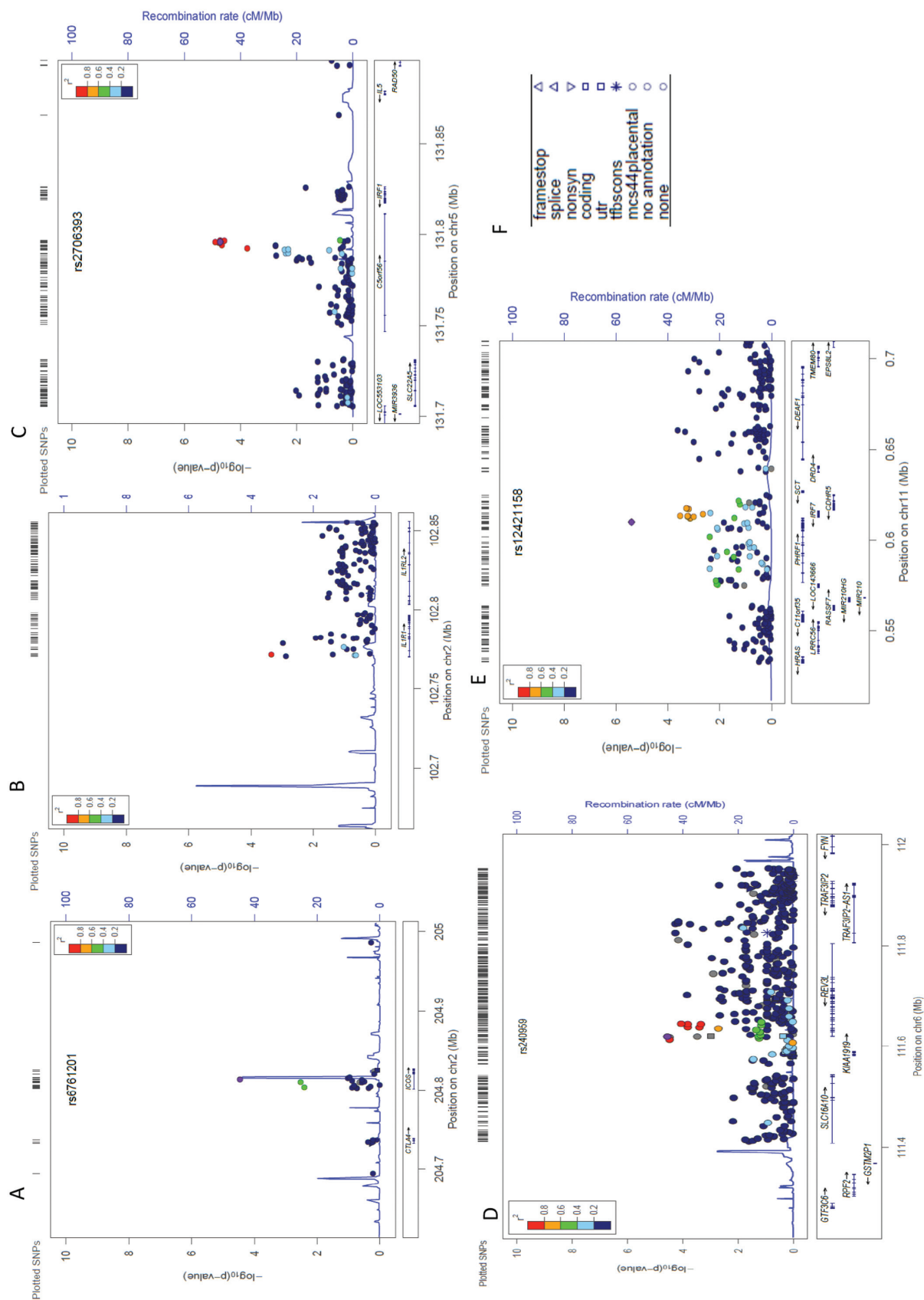


Figure 3. LocusZoom plots for non-HLA regions associated with RA in black South Africans.

tronic region of *ICOS* and *KIAA1542*. Furthermore, this study showed that variants of the *PTPN22* gene do not confer risk for RA in black South Africans. This is expected, as the MAF of the lead SNP in this gene is very low in the black population (10).

Although none of the significantly associated SNPs in *HLA-DRB1* locate to the coding region, the effect over the HLA region is very likely due to *HLA-DRB1* alleles. Using conditional analysis, this study was the first to demonstrate that *HLA DRB1* completely explains the risk for RA in the extended MHC. Unlike in Caucasians, we found no associations with *HLA B* or *HLA DPB1* (34–36), although our sample has limited statistical power to detect these secondary effects in the MHC.

Numerous HLA-associated SNPs locating to the intergenic regions *BTNL2* | *HLA-DRA*, *LOC442175* | *ZNF165*, *PSMB9* | *HLA-DMB* and to the coding region of *CCHCR1* reached genome-wide significance. However, conditional analysis revealed that the effect can largely be explained by the strong linkage disequilibrium (LD) with *HLA-DRB1*.

The association of two SNPs close to the *RBPJ* gene on chromosome 4, rs874040 and rs36110812, in the intergenic region *LOC389203* | *RBPJ*, are of interest. The former SNP was previously found to confer a very modest risk (OR = 1.14) for RA in Caucasians (37). The *RBPJ* gene, which is essential for the Notch pathway, controls numerous cell-fate specification events. The protein encoded for by the *RBPJ* gene is a transcriptional regulator that binds specifically to the immunoglobulin kappa-type J segment recombination signal sequence and acts as both a transcriptional repressor and activator (38).

Associations with several SNPs in or near the interleukin 1 receptor, type 1 (*IL1R1*) and the interferon regulatory factor 1 (*IRF1*) gene were found. The former is part of the toll-like receptor superfamily and codes for receptors for interleukin-1 α (IL-1 α), interleukin-1 β (IL-1 β), and interleukin-1 receptor an-

tagonist (IL-1RA). This receptor interacts with molecules such as MyD88, IRAK1, IRAK4 and TRAF6. Variants of this gene have been associated with asthma (39) and with severe hand osteoarthritis (40). Numerous significantly associated SNPs locate to the *IRF1* gene on chromosome 5, which is responsible for the activation of interferon α and β . Knockout mice with deletion of *IRF1* had abnormal peripheral blood lymphocytes, specifically decreased CD8⁺-positive T cell and natural killer (NK) cell numbers and an increase in CD4⁺-positive T cells (41). Although none of these SNPs have been associated with RA in other studies, rs2522056 has been associated with Crohn's disease (42), an increase in acute phase response (43) and fibrinogen levels and, therefore, can be considered a risk factor for cardiovascular disease (44).

The *TRAF3IP2* gene has been associated with the risk of developing psoriasis (45) and psoriatic arthritis (46); *ICOS*, with alopecia areata; *KIAA1542*, with the presence of anti-dsDNA antibody positivity in systemic lupus erythematosus (47), with the function of the intergenic region between *KIAA1919* and *REV3L* being unknown.

None of the non-HLA loci previously associated with RA in Caucasians could be definitively shown to confer risk for RA in this study on black South Africans. The possible reason for the lack of association in this genetically distinct population may be that these loci are truly not risk loci in the studied population. However, this study was underpowered to detect the modest-to-small effects of many of these loci. In addition, the data show that some of the risk variants in Caucasians are nonpolymorphic in black South Africans, and in others the MAF is significantly lower. For example, the well-studied variant rs2476601, which encodes an amino acid change (R620W) in one of four SH3 domain binding sites in the *PTPN22* molecule, is monomorphic in this population. Overall, the corroboration of variants associated with RA in Caucasians with association in this popu-

lation is similar to the findings in the admixed African-American population (48).

Despite the smaller sample size of the seropositive subgroups (aCCP and RF), the strength of association of the HLA and non-HLA SNPs were very similar to the overall cohort, suggesting that the subgroups represent a more genetically homogenous group.

The ancestry informative markers and principle component analyses (PCA) showed a distinction of the black South African RA cases and controls from Caucasians (CEU), West African populations, that is, Yoruba of Nigeria (YRI) and the East Africans, the Luhya (LWK) and Maa-sai (MKK) tribes of Kenya (31). This genetic diversity between Africans is supported by earlier studies of the frequency of carriers of the shared epitope (SE) alleles. In black South Africans, more than 90% of RA cases carry at least one copy of the SE alleles (27) compared with a much lower frequency of 30% in a Cameroonian population in West Africa (28).

Using a high-throughput genotyping platform such as the ImmunoChip array allowed for the identification of admixed individuals and thus for better quality control of the dataset where population structure as a confounder could be avoided. This study further highlights that black South Africans are genetically distinct from populations resident in the East and West Africa. These findings emphasize the need for a unique reference set of data for Southern Africans, which was not available at the start of this study. Understanding the genomic architecture of the South African population will allow better study designs considering LD in this population.

CONCLUSION

The overall significance of this study is that it has given insight into the genetics of RA in black South Africans. There are risk loci that have been shown to be shared among all populations and others that are specific to this study population.

The ImmunoChip was designed on the basis of Caucasian GWAS data and is therefore not ideally suited for genotyp-

ing other ethnic groups. Africans differ from Caucasians and Asians in terms of their low LD structure and therefore many more tagging SNPs are required to cover the African genome (49). Current genotyping platforms have inadequate cover of variation in African genomes. The relatively small sample size meant that there was inadequate power to detect the modest effects that some non-HLA-associated loci might confer on susceptibility for RA in black South Africans.

This study did not address other causes of missing heritability such as gene–gene interaction, gene–environment interaction and epigenetic factors. Gene–environment interactions may contribute to the differences in the genetic risk for RA between Caucasians and black South Africans. Differences in the prevalence of RA have been reported between rural and urban black South Africans (50) and this suggests that urban black South Africans are exposed to some environmental risk factors that rural blacks are not.

Replication of the newly identified RA-associated loci is necessary. There is a clear need to conduct genetic studies in a larger sample size of black South Africans with RA to identify risk variants with smaller effect sizes. Furthermore a clinical phenotype has been described in Africans, “the African variant” which predominately affects larger joints and spares the smaller joints of the hand, unlike classic RA observed in Europeans (51). Performing studies in this subgroup will shed light on this unique phenotype with the hope of identifying pathogenic pathways that may be used to design individualized, targeted therapy.

ACKNOWLEDGMENTS

This study was made possible by a grant to N Govind from Carnegie Corporation of New York, New York, NY, USA (B8749). The authors would like to acknowledge the Connective Tissue Diseases Fund, University of the Witwatersrand, Johannesburg, South Africa, and the Medical Research Council of South Africa for financial support. SL Bridges Jr

acknowledges NIH grant R01 AR057202. A Choudhury acknowledges postdoctoral fellowships from National Research Foundation, South Africa, and SPARC postdoctoral fellowship program, University of the Witwatersrand, for financial support. RJ Reynolds is supported by NIH-K01AR060848. S Hazelhurst and M Ramsay acknowledge financial support from the National Research Foundation, South Africa.

DISCLOSURES

The authors declare they have no competing interests as defined by *Molecular Medicine*, or other interests that might be perceived to influence the results and discussion reported in this paper.

REFERENCES

- MacGregor AJ, et al. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* 43:30–7.
- Knight JC. (2013) Genomic modulators of the immune response. *Trends Genet.* 29:74–83.
- Trynka G, et al. (2011) Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* 43:1193–201.
- Juran BD, et al. (2012) Immuno-chip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. *Hum. Mol. Genet.* 21:5209–21.
- Beecham AH, et al. (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* 45:1353–60.
- Cooper JD, et al. (2012) Seven newly identified loci for autoimmune thyroid disease. *Hum. Mol. Genet.* 21:5202–8.
- Liu JZ, et al. (2012) Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat. Genet.* 45:670–5.
- Hinks A, et al. (2013) Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat. Genet.* 45:664–9.
- Eyre S, et al. (2012) High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* 44:1336–40.
- 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.
- Plenge RM, et al. (2005) Replication of putative candidate-gene associations with rheumatoid

- arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. *Am. J. Hum. Genet.* 77:1044–60.
- Lee HS, et al. (2009) Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations. *Arthritis Rheum.* 60:364–71.
- Ikari K, et al. (2006) Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population. *Rheumatology (Oxford).* 45:1345–8.
- Tikly M, Govind N, Frost J, Ramsay M. (2010) The PTPN22 R620W polymorphism is not associated with systemic rheumatic diseases in South Africans. *Rheumatology (Oxford).* 49:820–1.
- Ikari K, et al. (2005) Association between PADI4 and rheumatoid arthritis: a replication study. *Arthritis Rheum.* 52:3054–7.
- Kang CP, Lee HS, Ju H, Cho H, Kang C, Bae SC. (2006) A functional haplotype of the PADI4 gene associated with increased rheumatoid arthritis susceptibility in Koreans. *Arthritis Rheum.* 54:90–6.
- Suzuki A, et al. (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidyl-larginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* 34:395–402.
- Barton A, et al. (2004) A functional haplotype of the PADI4 gene associated with rheumatoid arthritis in a Japanese population is not associated in a United Kingdom population. *Arthritis Rheum.* 50:1117–21.
- Burr ML, et al. (2010) PADI4 genotype is not associated with rheumatoid arthritis in a large UK Caucasian population. *Ann. Rheum. Dis.* 69:666–70.
- Caponi L, et al. (2005) A family based study shows no association between rheumatoid arthritis and the PADI4 gene in a white French population. *Ann. Rheum. Dis.* 64:587–93.
- Martinez A, et al. (2005) PADI4 polymorphisms are not associated with rheumatoid arthritis in the Spanish population. *Rheumatology (Oxford).* 44:1263–6.
- Remmers EF, et al. (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.* 357:977–86.
- Lee HS, Remmers EF, Le JM, Kastner DL, Bae SC, Gregersen PK. (2007) Association of STAT4 with rheumatoid arthritis in the Korean population. *Mol. Med.* 13:455–60.
- Martell RW, du Toit ED, Kalla AA, Meyers OL. (1989) Association of rheumatoid arthritis with HLA in three South African populations—whites, blacks and a population of mixed ancestry. *S. Afr. Med. J.* 76:189–90.
- Mody GM, Hammond MG, Naidoo PD. (1989) HLA associations with rheumatoid arthritis in African blacks. *J. Rheumatol.* 16:1326–8.
- Pile KD, Tikly M, Bell JI, Wordsworth BP. (1992) HLA-DR antigens and rheumatoid arthritis in black South Africans: a study of ethnic groups. *Tissue Antigens.* 39:138–40.
- Meyer PW, et al. (2011) HLA-DRB1 shared epitope genotyping using the revised classification

- and its association with circulating autoantibodies, acute phase reactants, cytokines and clinical indices of disease activity in a cohort of South African rheumatoid arthritis patients. *Arthritis Res. Ther.* 13:R160.
28. Singwe-Ngandeu M, Finckh A, Bas S, Tiercy JM, Gabay C. (2010) Diagnostic value of anti-cyclic citrullinated peptides and association with HLA-DRB1 shared epitope alleles in African rheumatoid arthritis patients. *Arthritis Res. Ther.* 12:R36.
 29. Arnett FC, *et al.* (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* 31:315–24.
 30. Cano P, *et al.* (2007) Common and well-documented HLA alleles: report of the Ad-Hoc committee of the American Society for Histocompatibility and Immunogenetics. *Hum. Immunol.* 68:392–417.
 31. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. (2010) Data quality control in genetic case-control association studies. *Nat. Protoc.* 5:1564–73.
 32. 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.
 33. May A, *et al.* (2013) Genetic diversity in black South Africans from Soweto. *BMC Genomics.* 14:644.
 34. Lee HS, *et al.* (2008) Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus. *Mol. Med.* 14:293–300.
 35. Raychaudhuri S, *et al.* (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* 44:291–6.
 36. Ding B, *et al.* (2009) Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. *Arthritis Rheum.* 60:30–8.
 37. Stahl EA, *et al.* (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* 42:508–14.
 38. Castel D, Mourikis P, Bartels SJ, Brinkman AB, Tajbakhsh S, Stunnenberg HG. (2013) Dynamic binding of RBPJ is determined by Notch signaling status. *Genes. Dev.* 27:1059–71.
 39. Grotenboer NS, Ketelaar ME, Koppelman GH, Nawijn MC. (2013) Decoding asthma: translating genetic variation in IL33 and IL1RL1 into disease pathophysiology. *J. Allergy Clin. Immunol.* 131:856–65.
 40. Nakki A, *et al.* (2010) Allelic variants of IL1R1 gene associate with severe hand osteoarthritis. *BMC Med. Genet.* 11:50.
 41. Gardin A, White J. (2011) The Sanger Mouse Genetics Programme: high throughput characterisation of knockout mice. *Acta Ophthalmol.* 89 Suppl s248.
 42. Franke A, *et al.* (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* 42:1118–25.
 43. Dehghan A, *et al.* (2011) Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation.* 123:731–8.
 44. Dehghan A, *et al.* Association of novel genetic Loci with circulating fibrinogen levels: a genome-wide association study in 6 population-based cohorts. *Circ. Cardiovasc. Genet.* 22:125–33.
 45. Ellinghaus E, *et al.* (2010) Genome-wide association study identifies a psoriasis susceptibility locus at TRAF3IP2. *Nat. Genet.* 42:991–5.
 46. Huffmeier U, *et al.* (2010) Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis. *Nat. Genet.* 42:996–9.
 47. Chung SA, *et al.* (2011) Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PLoS. Genet.* 7:e1001323.
 48. Hughes LB, *et al.* (2010) Most common SNPs associated with rheumatoid arthritis in subjects of European ancestry confer risk of rheumatoid arthritis in African-Americans. *Arthritis Rheum.* 62:3547–53.
 49. Teo YY, Small KS, Kwiatkowski DP. (2010) Methodological challenges of genome-wide association analysis in Africa. *Nat. Rev. Genet.* 11:149–60.
 50. Beighton P, Solomon L, Valkenburg HA. (1975) Rheumatoid arthritis in a rural South African Negro population. *Ann. Rheum. Dis.* 34:136–41.
 51. Maritz NG, Gerber AJ, Greyling SJ, Sanda BB. (2003) The rheumatoid wrist in black South African patients. *J. Hand Surg. Br.* 28:373–5.