Identification of the **NF-κB** activating protein-like locus as a risk locus for rheumatoid arthritis

Gang Xie,1,3 Yue Lu,2 Ye Sun,1,3 Steven Shiyang Zhang,1 Edward Clark Keystone,3 Peter K Gregersen,4 Robert M Plenge,5 Christopher I Amos,2 Katherine A Siminovitch,1,3,6

**ABSTRACT**

**Objective** To fine-map the NF-κB activating protein-like (**NKAPL**) locus identified in a prior genome-wide study as a possible rheumatoid arthritis (**RA**) risk locus and thereby delineate additional variants with stronger and/or independent disease association.

**Methods** Genotypes for 101 single nucleotide polymorphisms (SNPs) in the **NKAPL** locus on chromosome 6p22.1 were obtained on 1368 Canadian RA cases and 1471 controls. Single marker associations were examined using logistic regression and the most strongly associated **NKAPL** locus SNPs were then typed in another Canadian and a US-based RA case/control cohort.

**Results** Fine-mapping analyses identified six **NKAPL** locus variants in a single haplotype block showing association with p≤5.6×10−8 in the combined Canadian cohort. Among these SNPs, rs35656932 in the zinc finger F63 gene and rs13208096 in the **NKAPL** gene remained significant after conditional logistic regression, were conditionally independent for risk, and were replicated in the US cohort (P<0.0001 and 2.44×10−9, respectively). These associations remained significant after conditioning on SNPs tagging the HLA-DRB1 locus.

**Conclusions** By illuminating additional **NKAPL** variants with high strength in the HLA-DRB1 locus, our data conclusively identify **NKAPL** as an RA susceptibility locus.

**INTRODUCTION**

Rheumatoid arthritis (**RA**) is a chronic autoimmune disease primarily associated with inflammation of the synovial joints and affecting up to 1% of the population worldwide. Although the complex interplay of genetic and environmental factors underpinning RA are not well understood, major inroads have been made in mapping gene loci associated with risk for this disease.1 In addition to the **HLA-DRB1** locus, over 55 non-major histocompatibility complex (**MHC**) RA risk loci have emerged from genome-wide association studies (**GWAS**) and subsequent GWAS meta-analyses of the GWAS datasets.2–10

Through a genome-wide scan of 2418 **RA** patients and 4504 healthy controls ascertained in Canada and the USA, we previously identified an association of RA with the REL **NF-κB** transcription factor locus and also confirmed already identified disease associations with the **PTPN22**, **CTLA4**, **TNFAIP3**, **BLK** and **TRA F4/C5** genes.9 Our data also showed strongly suggestive signals (**PGW AS** values between 8.2×10−7 and 5.28×10−6) emanating from a cluster of Single nucleotide polymorphisms (**SNP**) across a 70 kb region on chromosome 6p22.1 encompassing the **NF-κB** activating protein-like gene (**NKAPL**) as well as three Zinc finger protein transcription factors **ZNF613**, **ZNF307** and **ZNF687**. A follow-up GWAS meta-analysis of 5505 **RA** and 22603 controls of European descent10 also revealed 15 genotyped or imputed SNPs across a 150 kb region encompassing the **NKAPL** locus to be strongly associated (**P meta** between 1×10−10 and 1×10−15) with RA (R. Plenge, personal communication). Although **NKAPL** functions are unknown, its protein product shows 90% sequence similarity to **NF-κB** activating protein (**NKAP**), a protein implicated in **NF-κB**-mediated transcriptional activation of **TNF** and IL-1.11 As data from our group and others have also implicated other genes from the **NF-κB** signalling pathway (eg, **REL**, **CD40**, **TRA F4**, **TNFAIP3**, **PR KCQ** and **TNFRSF4**) in **RA** susceptibility,12 13 the **NKAPL** gene represents a compelling potential candidate gene for RA. We therefore undertook fine-mapping studies of the **NKAPL** locus aimed at confirming this association, identifying those variants providing the strongest association signal and defining whether such variants act together or independently of one another and/or the **HLA-DRB1** locus in conferring risk for RA.

**METHODS**

**Subjects** Study cohorts (see online supplementary methods) include: 5979 subjects of European origin (2078 **RA** patients and 1901 healthy controls) recruited independently from two clinical centres in Canada, Toronto (1568 cases and 1471 healthy controls) and Halifax (710 cases and 450 healthy controls) and a third cohort including 2064 subjects of European ancestry ascertained in the USA as part of the Brigham Rheumatoid Arthritis Sequential Study and used here for replication analysis.

**SNP selection** SNPs from a 372 kb interval across the **NKAPL** locus on chromosome 6p22.1 were selected primarily based on at least one of the following criteria:...
(1) HapMap phase III data identifying the SNP as a tag SNP with minor allele frequency >0.01 and $r^2$ threshold of 0.8 or (2) localisation within 150 kb upstream or downstream of SNPs most significantly associated with RA in our GWAS. Other SNPs studied were: the autoimmune disease-associated PTBP2 rs2476601\(^{14}\) and two SNPs (rs660895 and rs6910071) that tag the *HLA-DRB1*\(^{14}\) allele on chromosome 6p21.3.\(^{15}\)

**Statistical analyses**

Hardy–Weinberg equilibrium, allelic association and conditional logistic regression analyses were performed using PLINK software V1.07 (http://pngu.mgh.harvard.edu/purcell/plink/). For the allelic association tests, the threshold for declaring significance was assigned according to Benjamini and Hochberg’s False Discovery Rate method and set at $p<5.00\times10^{-4}$ (0.05/101). Cochran–Mantel Haenszel $\chi^2$ method was used to calculate OR from the Canadian and US cohorts and risk alleles, respectively. For multinomial logistic regression analyses of multiple markers were performed using SAS V9.13 (SAS Institute Inc., Cary, North Carolina, USA). Gene–gene interaction analysis was performed by case-only interaction analysis in which a logistic regression model was used to test for an association of shared epitope (SE) positivity with NFKAPL risk alleles (coded in an additive fashion as –1, 0 or 1 for no, 1 or 2 risk alleles, respectively). For multinomial logistic regression modelling,\(^{17}\) controls were considered as the lowest risk outcome, SE negative cases as the intermediate risk outcome and SE positive cases as the highest risk outcome, and these multiple outcomes were then assessed according to number of NFKAPL risk alleles. The statistical power for this study was evaluated using CaTS software (http://www.sph.umich.edu/cgi/abecasis/CaTS/) with the following parameters: disease prevalence 0.01, disease allele frequency 0.2, $\alpha=0.0005$ (0.05/101). Power to detect associations with relative risk of 1.5 was estimated to be 99.4%.

**RESULTS**

Fine-mapping of the RA-associated NFKAPL locus at 6p22.1

To identify risk allele(s) at the NFKAPL locus, we genotyped 1568 RA cases and 1471 controls from Toronto for 105 SNPs across a 372 kb genomic region encompassing the NFKAPL gene. Characteristics of the study design and subjects are outlined in online supplementary figure S1. Among the 101 SNPs that passed quality control, 16 achieved the set significance threshold of $p<5.00\times10^{-4}$ with the top six markers showing associations with disease ($p<6.00\times10^{-7}$) that remained highly significant (p values $1.80\times10^{-4}$–$6.40\times10^{-5}$) after False Discovery Rate correction (table 1 and online supplementary table S1).

Haplotype analysis of pairwise LD among the 101 SNPs revealed that these six most strongly associated SNPs map within a 70 kb segment representing the middle of three haplotype blocks across this region and containing the NFKAPL gene and three zinc finger transcription factor genes, ZNF193, ZNF307 and ZNF187 (figure 1 and see online supplementary figure S2). The strongest association signal ($p=2.48\times10^{-7}$) came from a ZNF187 intronic SNP (rs67998226) at the distal end of this haplotype block, but these variants were all in strong LD with one another, the pairwise LD

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Risk allele</th>
<th>Sample cohort</th>
<th>OR (95% CI)</th>
<th>$p$ value*</th>
<th>$P_{\text{FDR}}$ value</th>
<th>Cochran Q $p$ value†</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs13195291</td>
<td>ZNF193</td>
<td>A</td>
<td>Toronto</td>
<td>1.61 (1.33 to 1.94)</td>
<td>$5.03\times10^{-7}$</td>
<td>$8.60\times10^{-6}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Halifax</td>
<td>1.40 (1.04 to 1.89)</td>
<td>0.03</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Combined‡</td>
<td>1.56 (1.33 to 1.82)</td>
<td>$2.79\times10^{-8}$</td>
<td>$4.20\times10^{-8}$</td>
<td>0.45</td>
</tr>
<tr>
<td>rs35656832</td>
<td>ZNF193</td>
<td>T</td>
<td>Toronto</td>
<td>1.66 (1.38 to 1.99)</td>
<td>$3.82\times10^{-8}$</td>
<td>$1.80\times10^{-8}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Halifax</td>
<td>1.48 (1.11 to 1.99)</td>
<td>$8.31\times10^{-3}$</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>1.62 (1.39 to 1.89)</td>
<td>$7.34\times10^{-10}$</td>
<td>$2.20\times10^{-9}$</td>
<td>0.52</td>
</tr>
<tr>
<td>rs13204012</td>
<td>ZNF193</td>
<td>A</td>
<td>Toronto</td>
<td>1.63 (1.35 to 1.95)</td>
<td>$1.53\times10^{-7}$</td>
<td>$3.90\times10^{-6}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Halifax</td>
<td>1.29 (0.96 to 1.74)</td>
<td>0.09</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>1.53 (1.31 to 1.79)</td>
<td>$5.66\times10^{-8}$</td>
<td>$5.60\times10^{-8}$</td>
<td>0.20</td>
</tr>
<tr>
<td>rs17720293</td>
<td>ZNF307</td>
<td>T</td>
<td>Toronto</td>
<td>1.60 (1.35 to 1.90)</td>
<td>$5.27\times10^{-8}$</td>
<td>$1.80\times10^{-6}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Halifax</td>
<td>1.23 (0.94 to 1.60)</td>
<td>0.139</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>1.49 (1.29 to 1.72)</td>
<td>$4.29\times10^{-8}$</td>
<td>$5.10\times10^{-8}$</td>
<td>0.10</td>
</tr>
<tr>
<td>rs13208096</td>
<td>NFKAPL</td>
<td>G</td>
<td>Toronto</td>
<td>1.67 (1.34 to 1.94)</td>
<td>$3.69\times10^{-7}$</td>
<td>$7.50\times10^{-6}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Halifax</td>
<td>1.48 (1.10 to 1.99)</td>
<td>$8.75\times10^{-3}$</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>1.58 (1.35 to 1.85)</td>
<td>$6.99\times10^{-9}$</td>
<td>$1.40\times10^{-8}$</td>
<td>0.64</td>
</tr>
<tr>
<td>rs67998226</td>
<td>ZNF187</td>
<td>C</td>
<td>Toronto</td>
<td>1.66 (1.39 to 1.99)</td>
<td>$2.48\times10^{-8}$</td>
<td>$1.80\times10^{-6}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Halifax</td>
<td>1.47 (1.10 to 1.97)</td>
<td>$9.88\times10^{-3}$</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>1.62 (1.39 to 1.88)</td>
<td>$6.53\times10^{-10}$</td>
<td>$2.20\times10^{-9}$</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Genotype data are shown for six SNPs tested in 1368 cases and 1471 controls from Toronto in step 1 and replication in 710 cases and 430 controls from Halifax in step 2. Data for all other SNPs tested in Toronto cohort are shown in online supplementary table S1.

*Nominal $p$ value from an allele-based case-control comparison with 1 degree of freedom; $p<0.05$ are highlighted in bold type. $P_{\text{FDR}}$ corresponds to $p$ values adjusted for multiple testing using the False Discovery Rate Method implemented in PLINK.

†Genotyping data from the Toronto and Halifax cohorts were merged in PLINK and combined $P_{\text{FDR}}$ values for association evaluated using Cochran–Mantel Haenszel tests and the Cochran Q test for heterogeneity.
between rs67998226 and either the three SNPs across the more proximal ZNF193 gene (rs13195291, rs35656932 and rs13204012) or the NKAPL promoter region SNP (rs13208096) being >0.94 and 0.83, respectively, and between rs35656932 and rs13208096 being 0.87.

To further examine effects of this locus on RA susceptibility, the six most significant SNPs were also typed in a Halifax-derived cohort including 710 RA patients and 430 controls. Four of the six associations (rs13195291, rs35656932, rs13208096 and rs67998226) were replicated in this cohort, with combined (Toronto and Halifax) association signals (PCAN) ranging from 5.60×10⁻⁸ to 2.22×10⁻⁹ (table 1).

The extent to which the signals observed in the combined analysis were independent of one another was next examined using stepwise logistic regression analysis wherein variables were iteratively added into an empty model. This analysis identified the ZNF193/NKAPL locus rs55656932 SNP, which is in strong LD (r²=0.96) with the ZNF187 rs67998226 SNP as the variant most strongly associated with risk for RA, but also suggested that both NKAPL rs13208096 and rs35656932 SNP alleles influence risk for RA (table 2). Additional conditional analyses of the six SNPs revealed that both rs35656932 and rs13208096 remain significantly associated with disease after conditioning on each marker (see online supplementary table S2). Because these data suggest independent contributions of alleles of rs55656932 and rs13208096 on risk, associations of these two variants with disease were next explored in a third (US) cohort including 863 cases and 1201 controls. The two SNPs both
repli- cated in this cohort (table 3): combined analyses of all 2941 cases and 3102 controls typed for these SNPs (table 3) yielding high- ly significant $P_{\text{comb}}$ values ($4.24 \times 10^{-16}$ for rs35656932 and $2.44 \times 10^{-9}$ for rs15208096).

**Table 2** Stepwise logistic regression in the two SNP models

<table>
<thead>
<tr>
<th>SNP added to the model</th>
<th>p Value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs35656932</td>
<td>0.003</td>
<td>5.82 (1.82 to 18.58)</td>
</tr>
<tr>
<td>rs13208096</td>
<td>0.030</td>
<td>3.57 (1.12 to 11.11)</td>
</tr>
</tbody>
</table>

Stepwise logistic regression results based on analysis of the six most strongly associated NKAPL locus SNPs in the 2078 RA cases and 1901 controls from Canada.

**Variants of the NKAPL locus and the HLA region jointly contribute to risk for RA**

Because the NKAPL locus maps to chromosome 6p22.1, evaluation of its disease association may be confounded by effects of the HLA-DRB1 locus at 6p21.3, a locus which encodes the RA-predisposing SE, confers much of the genetic risk for RA, and maps in a region of extensive LD. Although the NKAPL gene lies 4306 kb upstream of the RA risk-related HLA-DRB1 gene, the extent to which its association with RA may reflect LD with HLA-DRB1 risk alleles was investigated by genotyping the Canadian cases (2078) and controls (1901) for two SNPs (rs6060895 and rs6910071) that tag one of the most common SE encoding alleles, HLA-DRB1*0401, re-assessing the NKAPL-disease association by logistic regression analysis with conditioning on the HLA-DRB1*0401 SNPs. This analysis confirmed the strong association of RA with both tag SNPs, the signals reaching $p=2.04 \times 10^{-7}$ and $p=2.59 \times 10^{-6}$ for rs6060895 and rs6910071, respectively (see online supplementary table S5). Importantly, the association signals from each of the six NKAPL locus SNPs remained highly significant ($p<1 \times 10^{-10}$) after conditioning on the DRB1 tag SNPs (see online supplementary table S4). Analyses of the pairwise LD between the two tag SNPs and each of the 55 SNPs with nominal evidence ($p<0.05$) for disease association in the initial analysis of Toronto controls also revealed no significant LD ($r^2<0.01$ for each SNP pair) between any of these SNPs and the HLA-DRB1 tag SNPs (data not shown). These findings imply that the association signal at the NKAPL locus represents an effect on risk that is not attributable to LD with the HLA-DRB1*0401 allele.

To further evaluate the relationship between NKAPL and HLA locus effects on risk, we additionally used data from a large panel of British subjects genotyped for HLA alleles to impute, in the subset of the Canadian cohort included in our previous GWAS, 157 HLA alleles encoding classic HLA-A, B, C, DRB1, DQA and DQB molecules. Association tests were performed in a dataset combining the imputed alleles, and the GWAS-derived HLA region as well as the fine mapping-derived SNP genotypes. Single SNP association tests performed using this combined dataset and assuming an additive model (implemented in PLINK software) identified 579 SNPs (data not shown) or HLA genotypes with $p$ values less than $1.0 \times 10^{-4}$ (see online supplementary table S5). These variants together with the six candidate SNPs identified by fine mapping were subjected to stepwise logistic regression analysis using SAS and a $p$ value set at $<0.01$ as the criterion to enter and remain in the model. From this analysis, seven SNPs (including rs35656932), but no classical HLA alleles were retained in the model, suggesting that risk for disease at this locus is better explained by effects of SNPs rather than HLA alleles (see online supplementary table S6A). Stepwise logistic regression was then repeated with HLA-DRB1*0401 forced into the model. After this analysis, five variants, including HLA-DRB1*0401 (p = 0.04), were retained in the model, although this HLA allele did not reach our significance criterion of p < 0.01 (see online supplementary table S6B). These data provide further evidence that NKAPL variant(s) per se contribute to risk for RA.

Relevance of HLA genotypes to the NKAPL locus effect on risk for RA was also explored by assessing the extent to which the NKAPL region rs35656932 and rs13208096 SNPs associate with disease in cases stratified based on presence or absence of the SE alleles or of anticyclic citrullinated peptide (anti-CCP) antibody, an autoantibody strongly associated with SE alleles. These data provide further evidence that NKAPL variant(s) per se contribute to risk for RA.

**Table 3** Associations of the NKAPL locus rs13208096 and rs13708096 SNPs with rheumatoid arthritis in the combined Canadian and US cohort

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Risk allele</th>
<th>Sample cohort</th>
<th>RAF</th>
<th>Case OR (95% CI)</th>
<th>$P_{\text{comb}}$</th>
<th>Meta Q p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs35656932</td>
<td>ZNF193</td>
<td>T</td>
<td>Toronto</td>
<td>0.118</td>
<td>0.075</td>
<td>1.66 (1.38 to 1.99)</td>
<td>3.82 $\times 10^{-8}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Halifax</td>
<td>0.119</td>
<td>0.083</td>
<td>1.48 (1.11 to 1.99)</td>
<td>8.31 $\times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>USA</td>
<td>0.088</td>
<td>0.070</td>
<td>1.29 (1.02 to 1.63)</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Combined*</td>
<td>0.109</td>
<td>0.074</td>
<td>1.50 (1.32 to 1.71)</td>
<td>4.24 $\times 10^{-10}$</td>
</tr>
<tr>
<td>rs13208096</td>
<td>NKAPL</td>
<td>G</td>
<td>Toronto</td>
<td>0.115</td>
<td>0.075</td>
<td>1.61 (1.34 to 1.94)</td>
<td>3.69 $\times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Halifax</td>
<td>0.116</td>
<td>0.081</td>
<td>1.48 (1.10 to 1.99)</td>
<td>8.75 $\times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>USA</td>
<td>0.086</td>
<td>0.068</td>
<td>1.30 (1.03 to 1.63)</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>0.107</td>
<td>0.073</td>
<td>1.48 (1.32 to 1.69)</td>
<td>2.44 $\times 10^{-9}$</td>
</tr>
</tbody>
</table>

*Combined $p$ values for association ($P_{\text{comb}}$) were performed using Cochran-Mantel Haenszel tests stratifying by centre. The Cochran Q tests for heterogeneity in the effect among studies.

NKAPL, NF-κB activating protein-like; RAF, risk allele frequency.
Effects of the NKAPL locus on risk were the strongest in the SE negative patients and increased when conditioning on HLA-DRB1 *0401 tag SNPs, suggesting interaction between NKAPL and the HLA-DRB1 risk alleles. This possibility was directly examined using case-only logistic regression models wherein SE positivity was considered the outcome and NKAPL genotype an additive effect. This analysis (table 5) revealed significance between risks conferred by HLA-DRB1 alleles and the NKAPL rs35656932 and rs13208096 SNPs, with the interaction effect being negative for both the latter SNPs (OR=0.67/CI 0.54 to 0.83 and OR=0.69/CI 0.55 to 0.86, respectively). By comparison, a positive interaction effect was again revealed the ORs associated with either of the two SNPs and the NKAPL rs35656932 and rs13208096 SNPs, with the interaction effect being negative for both the latter SNPs (OR=0.67/CI 0.54 to 0.83 and OR=0.69/CI 0.55 to 0.86, respectively). By comparison, a positive interaction effect was apparent between the RA-associated PTPN22 rs2476601 variant and SE positivity (OR=1.54, CI 1.08 to 1.65). Results of multinomial logistic regression analyses (see online supplementary table S7) further confirmed strong interaction of the HLA-DRB1 SE alleles with each of the NKAPL (p=4.55×10^-10 and 2.51×10^-11) and the PTPN22 (p=2.89×10^-15) risk alleles and again revealed the ORs associated with either of the two NKAPL risk alleles to be lower, but for the PTPN22 risk allele to be higher, in a comparison of SE positive cases to controls versus SE negative cases to controls.

**DISCUSSION**

The NKAPL region emerged as a candidate RA risk locus in the context of our prior GWAS data providing strong evidence for association of SNPs across this locus with RA in a Canadian and US study population. Because the association was supported in subsequent meta-analysis incorporating this and four more GWAS datasets, we undertook fine mapping and conditional analyses to screen for risk variants with stronger and/or independent signals of disease association. By genotyping our Canadian RA case/control cohort for 101 SNPs across the locus, we have identified six variants for which association with RA reaches a conservative level of genome-wide significance (p<5.7×10^-8). Haplotype analysis reveals these six variants to all lie within one of these haplotype blocks in a region encompassing the NKAPL and three zinc finger transcription factor genes. These genes are in strong LD with one another, but results of stepwise and conditional logistic regression analyses indicate that both ZNF195 rs35656932 and NKAPL rs13208096 SNPs contribute to risk for RA, associations for these two markers remaining significant after conditioning for each other associated SNP and results of stepwise logistic regression also suggesting independent effects of these SNPs on risk. These two associations were also replicated, albeit at modest levels of significance, in an independent US-based cohort.

Interpreting effects of the NKAPL locus on RA risk is complicated by the location of this locus in a chromosomal region (6p22.1) upstream of the HLA class II genes. While the NKAPL locus maps about 4306 kb away from the HLA-DRB1 gene and 1386 kb upstream from the telomeric end of the HLA region, the extensive LD across the region raises the possibility that the NKAPL association signals reflect LD with HLA-DRB1 SE alleles. However, effects of NKAPL as well as DRB1 SNPs on disease risk were revealed here by the logistic regression analyses conditioning on either of two HLA-DRB1*0401 tag SNPs. Analysis of pairwise LD between each of the most strongly associated NKAPL SNPs and the HLA-DRB1 tag SNPs also provided no evidence for LD (r^2<0.01) between SNPs at these respective loci and a stepwise logistic regression analysis combining the top six candidate SNPs from the fine-mapping study, 379 HLA region SNPs from the GWAS and the imputed HLA alleles further support contribution of the NKAPL locus to risk for RA. These findings are consistent with other data suggesting that the MHC locus contains loci in addition to HLA-DRB1 that confer risk to RA. Importantly, primary association of this locus with SE negative disease also implies interaction between the NKAPL and HLA-DRB1 risk alleles, a possibility supported by the results of case-only and multinomial logistic analyses showing very significant negative interaction effects.
between HLA-DRB1 and each of two NKAPl risk alleles. While both the NKAPl and the observed PTPN22-HLA-DRB1 interaction affects on risk require further evaluation in relation to their biological significance, these data provide new insights into the complex effector interactions that may link risk genotypes to disease pathogenesis in RA.

The current data identify two SNPs, rs13208096 and rs5656952, as the major drivers of the association signal at the NKAPl locus. Among these, rs13208096 maps 1787bp upstream of NKAPl gene expressed in many tissues, including most immune cell populations. The gene encodes a 402 amino acid nuclear protein highly homologous to the NKAPl that functions as a transcriptional repressor of NOTCH signalling in thymocytes and is required for haemopoietic stem cell maintenance and survival. NKAPl is not functionally characterised, but shares with NKAPl a domain critical to NKAPl roles in transcriptional repression and recent data from genome-wide annotation of transcriptional regulators (ENCONE:UCSC genome browser: http://genome.ucsc.edu/) reveal the rs13208096 SNP to be located in a region containing putative transcriptional regulatory histone marks.

The second SNP that drives the association signal at the NKAPl locus, rs5656952, maps 1741bp upstream of the ZNF193 gene encoding the zinc finger transcription protein, ZNF193. Little is known about the functions of ZNF193 or the ZNF187 locus, rs35656932, maps 1741bp upstream of the NKAPl locus. rs5656952 and ZNF507 proteins encoded by the nearby loci. However, ZNF507 has transcriptional repressor activity and appears to target NF-kB, raising the intriguing, albeit highly speculative possibility that several genes at the NKAPl locus influence RA risk via effects on NF-kB signalling.

Thus, while the current data do not identify the disease-causal allele at the NKAPl locus, our findings provide compelling evidence that this locus confers risk for RA and that the variants accounting for this association signal eminate from highly plausible candidate genes likely to influence NF-kB as well as HLA-DRB1-modulated immune cell responses already implicated in RA risk and pathogenesis.

Acknowledgements This work was supported by grants from the Canadian Institutes for Health Research (MOP74621) and Ontario Research Fund (RE-01-061). KAS holds the Sherman Family Chair in Genome Medicine and a Tier 1 Canada Research Chair. PAG, CIA and YL were partially supported by US NIH Grant AR444422.

Contributors GX planned and performed experiments, aided in writing the paper. Yue Lu and YS carried out statistical work, aided in preparing table. SSZ carried out DNA preparation and PCR reactions. ECK and RMP provided patient samples and data, aided in writing the paper. PKG provided data, aided in writing the paper. CIA prepared the paper, guided statistical analyses, aided in data interpretation and paper writing. KAS led the planning and performance of experiments, aided interpretation of data and writing of the paper.

Funding None.

Competing interests None.

Patient consent Obtained.

Ethics approval Mount Sinai Hospital Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non-Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc-3.0/.

REFERENCES


