EXTENDED REPORT

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]

► Additional data are published online only. To view these files please visit the journal online (http://dx.doi.org/10.1136/annrheumdis-2012-202076).

[1]Mount Sinai Hospital Samuel Lunenfeld Research Institute and Toronto General Research Institute, Toronto, Ontario, Canada
[2]Department of Epidemiology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA
[3]Rebecca MacDonald Centre for Arthritis, Department of Medicine, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada
[4]Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York, USA
[5]Division of Rheumatology, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts, USA
[6]Departments of Immunology and Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

Correspondence to
Dr Katherine A Siminovitch, Mount Sinai Hospital, Lunenfeld Research Institute and Toronto General Research Institute, 600 University Ave, Room 778D, Toronto, Ontario, Canada M5G 1X5; ksimin@mshri.on.ca

Received 23 May 2012
Accepted 21 October 2012
Published Online First 6 December 2012

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 SNPs across 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 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 $p \leq 5.6 \times 10^{-8}$ in the combined Canadian cohort. Among these SNPs, rs35656932 in the *zinc finger 193* gene and rs13208096 in the *NKAPL* gene remained significant after conditional logistic regression, contributed independently to risk for disease, and were replicated in the US cohort ($P_{comb} = 4.24 \times 10^{-10}$ and 2.44×10^{-9}, respectively). These associations remained significant after conditioning on SNPs tagging the HLA-shared epitope (SE) *DRB1*0401* allele and were significantly stronger in the HLA-SE negative versus positive subgroup, with a significant negative interaction apparent between *HLA-DRB1* SE and *NKAPL* risk alleles.

Conclusions By illuminating additional *NKAPL* variants with highly significant effects on risk that are distinct from, but interactive with those arising from 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-DRB* locus, over 35 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 *TRAF1/C5* genes.[9] Our data also showed strongly suggestive signals (P_{GWAS} values between 8.2×10^{-7} and 5.28×10^{-8}) 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 *ZNF193*, *ZNF307* and *ZNF187*. A follow-up GWAS meta-analysis of 5505 RA and 22 603 controls of European descent[10] also revealed 13 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^{-13}) 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, TRAF1, TNFAIP3, PRKCQ* and *TNFRSF14*) in RA susceptibility,[9 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: 3979 subjects of European origin (2078 RA patients and 1901 healthy controls) recruited independently from two clinical centres in Canada, Toronto (1368 cases and 1471 healthy controls) and Halifax (710 cases and 430 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 *PTPN22* rs2476601[14] and two SNPs (rs660895 and rs6910071) that tag the *HLA-DRB1*0401* 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 χ^2 analysis was used to combine p values and calculate OR from the Canadian and US cohorts and an R-script (http://www.rproject.org/) was used to generate figures. Haplotype block structure, depicted using Haploview software V4.1 (http://www.broad.mit.edu/mpg/haploview), was defined according to the criteria established by Gabriel[16] and the pairwise estimates of standardised Lewontin's disequilibrium coefficient (D′), whereas the linkage disequilibrium (LD) among pairs of SNPs was characterised according to the square of the correlation coefficient (r^2). Conditional 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 *NKAPL* 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 *NKAPL* risk alleles. The statistical power for this study was evaluated using CaTS software (http://www.sph.umich.edu/csg/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 *NKAPL* locus at 6p22.1

To identify risk allele(s) at the *NKAPL* locus, we genotyped 1368 RA cases and 1471 controls from Toronto for 105 SNPs across a 372 kb genomic region encompassing the *NKAPL* 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 1.80×10^{-6}–8.60×10^{-6}) after False Discovery Rate correction (table 1 and online supplementary table S1). Haploview 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 *NKAPL* gene and three zinc finger transcription factor genes, *ZNF193, ZNF307* and *ZNF187* (figure 1 and see online supplementary figure S2). The strongest association signal ($p=2.48\times10^{-8}$) 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 1 List of the six *NKAPL* locus SNPs showing the most significant association with rheumatoid arthritis in the Toronto- and Halifax-based case/control cohorts.

SNP	Gene	Risk allele	Sample cohort	RAF		OR (95% CI)	p Value*	P_{FDR} value	Cochran Q p value†
				Case	Control				
rs13195291	*ZNF193*	A	Toronto	0.115	0.075	1.61 (1.33 to 1.94)	**5.03×10^{-7}**	**8.60×10^{-6}**	
			Halifax	0.116	0.086	1.40 (1.04 to 1.89)	**0.03**	**0.04**	
			Combined†	0.116	0.077	1.56 (1.33 to 1.82)	**2.79×10^{-8}**	**4.20×10^{-8}**	0.45
rs35656932	*ZNF193*	T	Toronto	0.118	0.075	1.66 (1.38 to 1.99)	**3.82×10^{-8}**	**1.80×10^{-6}**	
			Halifax	0.119	0.083	1.48 (1.11 to 1.99)	**8.31×10^{-3}**	**0.02**	
			Combined	0.118	0.075	1.62 (1.39 to 1.89)	**7.34×10^{-10}**	**2.20×10^{-9}**	0.52
rs13204012	*ZNF193*	A	Toronto	0.116	0.075	1.63 (1.35 to 1.95)	**1.53×10^{-7}**	**3.90×10^{-6}**	
			Halifax	0.11	0.088	1.29 (0.96 to 1.74)	0.09	0.10	
			Combined	0.114	0.078	1.53 (1.31 to 1.79)	**5.66×10^{-8}**	**5.60×10^{-8}**	0.20
rs17720293	*ZNF307*	T	Toronto	0.135	0.089	1.60 (1.35 to 1.90)	**5.27×10^{-8}**	**1.80×10^{-6}**	
			Halifax	0.129	0.107	1.23 (0.94 to 1.60)	0.139	0.14	
			Combined	0.133	0.093	1.49 (1.29 to 1.72)	**4.29×10^{-8}**	**5.10×10^{-8}**	0.10
rs13208096	*NKAPL*	G	Toronto	0.115	0.075	1.61(1.34 to 1.94)	**3.69×10^{-7}**	**7.50×10^{-6}**	
			Halifax	0.116	0.081	1.48 (1.10 to 1.99)	**8.75×10^{-3}**	**0.02**	
			Combined	0.115	0.076	1.58 (1.35 to 1.85)	**6.99×10^{-9}**	**1.40×10^{-8}**	0.64
rs67998226	*ZNF187*	C	Toronto	0.122	0.077	1.66 (1.39 to 1.99)	**2.48×10^{-8}**	**1.80×10^{-6}**	
			Halifax	0.122	0.086	1.47 (1.10 to 1.97)	**9.88×10^{-3}**	**0.02**	
			Combined	0.122	0.079	1.62 (1.39 to 1.88)	**6.53×10^{-10}**	**2.20×10^{-9}**	0.48

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_{FDR} corresponds to p values adjusted for multiple testing using the False Discovery Method implemented in PLINK.

†Genotyping data from the Toronto and Halifax cohorts were merged in PLINK and combined P_{Can} values for association evaluated using Cochran-Mantel Haenszel tests and the Cochran Q test for heterogeneity.

FDR, False Discovery Rate; NKAPL, NF-κB activating protein-like; RAF, risk allele frequency.

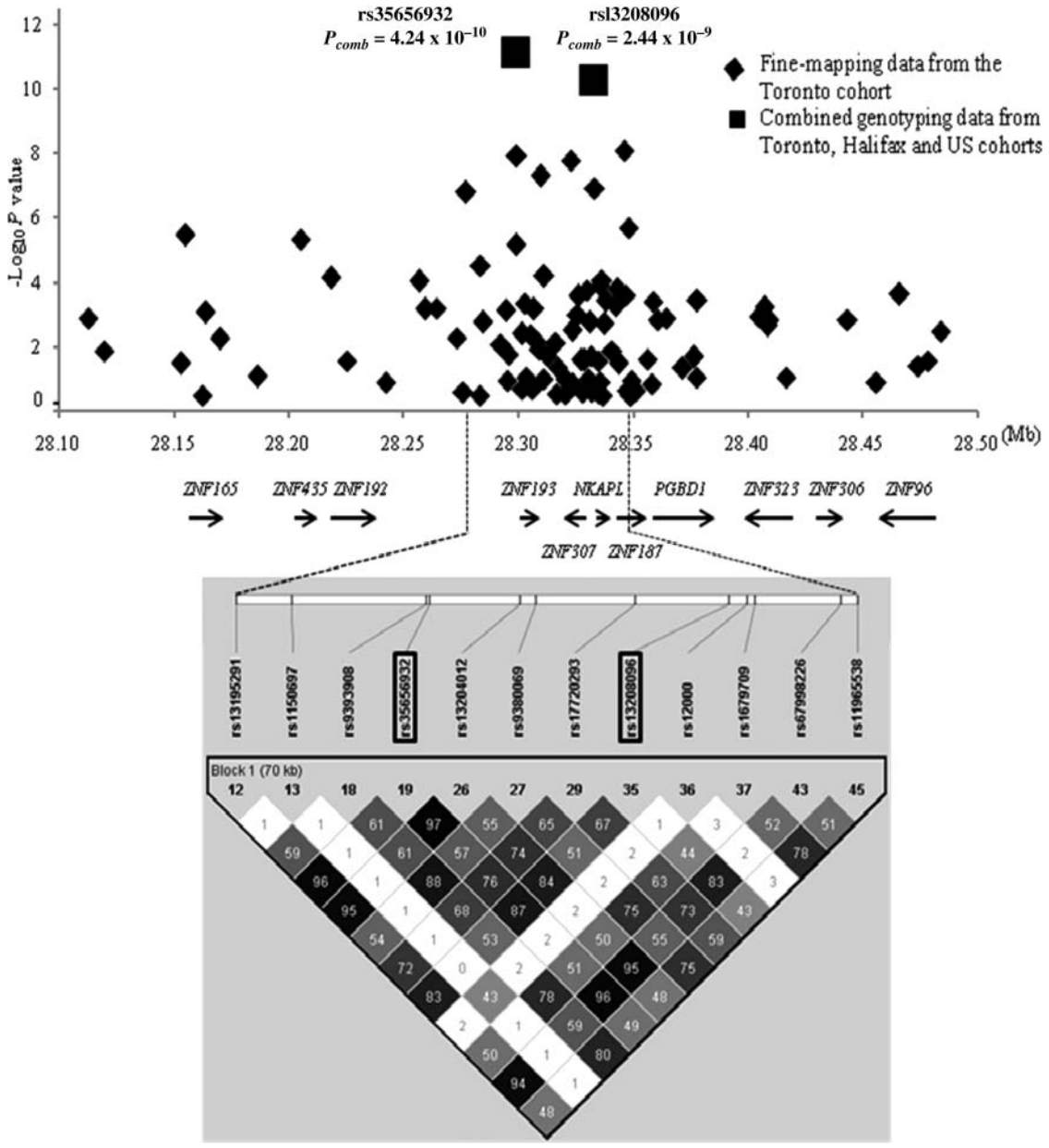

Figure 1 Case-control association results of SNPs (single nucleotide polymorphisms) at the NF-κB activating protein-like (NKAPL) locus on 6p22.1. SNPs genotyped in the fine mapping (◆) analysis of the Toronto-based case/control cohort. Results of combined analyses for the rs35656932 and rs13208096 are shown (■). Positions of the genes across the NKAPL locus are shown in the middle panel. The lower panel shows the linkage disequilibrium relationships between 12 of the most strongly associated SNPs ($p<5.0\times10^{-4}$) across an ~70 kb region encompassing the ZNF193, ZNF307, NKAPL and ZNF187 genes. The numbers shown in the squares in the lower panel are r^2 values representing the correlation coefficient for a given marker pair. The shade intensity lightens in parallel with diminishing correlation coefficient (r^2) values (r^2=1: black, r^2=0: white).

(r^2) 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 (P_{CAN}) 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 rs35656932 SNP, which is in strong LD (r^2=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 rs3656932 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 rs35656932 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

Table 2 Stepwise logistic regression in the two SNP models

SNP added to the model	p Value	OR (95% CI)
rs35656932	0.003	5.82 (1.82 to 18.58)
rs13208096	0.030	3.57 (1.12 to 11.11)

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.

replicated in this cohort (table 3): combined analyses of all 2941 cases and 3102 controls typed for these SNPs (table 3) yielding highly significant P_{comb} values (4.24×10^{-10} for rs35656932 and 2.44×10^{-9} for rs13208096).

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.[15 18–23] 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 (rs660895 and rs6910071) that tag one of the most common SE encoding alleles, *HLA-DRB1*0401*, reassessing 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 $p=2.04\times10^{-75}$ and $p=2.39\times10^{-63}$ for rs660895 and rs6910071, respectively (see online supplementary table S3). Importantly, the association signals from each of the six *NKAPL* locus SNPs remained highly significant ($p<1\times10^{-10}$) 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,[9] 137 *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 379 SNPs (data not shown) or HLA genotypes with p values less than 1.0×10^{-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.[24] This analysis revealed the disease associations for both risk alleles to be much higher in the SE negative ($p=1.20\times10^{-12}$ for rs35656932 and 3.50×10^{-11} for rs13208096) than in the SE positive ($p=4.70\times10^{-5}$ and 8.10×10^{-5}, respectively) subgroup (table 4). By contrast, the association signals from these loci were essentially the same in the anti-CCP positive and negative subsets. Because a disparity between SE and anti-CCP status effects on this association was not expected, the stratified subgroups were also genotyped for the *PTPN22* gene rs2476601 SNP for which an association with RA susceptibility is well established and thought to correlate with SE and anti-CCP positivity.[25–27] This analysis confirmed the strong association of RA with the rs2476601 variant ($p=2.72\times10^{-15}$; see online supplementary table S3) as well as the positive effects of the SE ($p=1.40\times10^{-16}$ SE positive vs 2.40×10^{-3} in SE negative cases)) and of anti-CCP antibody ($p=1.50\times10^{-11}$ in anti-CCP positive vs 5.10×10^{-2} in anti-CCP negative cases)) on this association. Thus, the PTPN22 risk allele is associated with SE positivity in an RA population in which NKAPL effects on risk are primarily observed in SE negative disease.

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

SNP	Gene	Risk allele	Sample cohort	RAF		OR (95% CI)	P_{Comb}	Meta Q p value
				Case	Control			
rs35656932	ZNF193	T	Toronto	0.118	0.075	1.66 (1.38 to 1.99)	3.82×10^{-8}	
			Halifax	0.119	0.083	1.48 (1.11 to 1.99)	8.31×10^{-3}	
			USA	0.088	0.070	1.29 (1.02 to 1.62)	0.032	
			Combined*	0.109	0.074	1.50 (1.32 to 1.71)	4.24×10^{-10}	0.23
rs13208096	NKAPL	G	Toronto	0.115	0.075	1.61 (1.34 to 1.94)	3.69×10^{-7}	
			Halifax	0.116	0.081	1.48 (1.10 to 1.99)	8.75×10^{-3}	
			USA	0.086	0.068	1.30 (1.03 to 1.63)	0.028	
			Combined	0.107	0.073	1.48 (1.32 to 1.69)	2.44×10^{-9}	0.35

*Combined p values for association (P_{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.

Table 4 Analysis of *NKAPL* and *PTPN22* risk allele disease associations in rheumatoid arthritis patient subgroups

SNP	Type	Cases	Controls	Risk allele	RAF		p Value*	OR (95% CI)
					Case	Control		
rs35656932	SE−	510	1901	T	0.150	0.077	1.20×10^{-12}	2.1 (1.7 to 2.6)
	SE+	1376	1901		0.105	0.077	4.70×10^{-5}	1.4 (1.2 to 1.7)
	Anti-CCP−	402	1901		0.128	0.077	1.70×10^{-6}	1.8 (1.4 to 2.3)
	Anti-CCP+	913	1901		0.112	0.077	3.00×10^{-6}	1.6 (1.3 to 1.9)
rs13208096	SE−	510	1901	G	0.145	0.076	3.50×10^{-11}	2.1 (1.7 to 2.6)
	SE+	1376	1901		0.105	0.076	8.10×10^{-5}	1.4 (1.2 to 1.7)
	Anti-CCP−	402	1901		0.128	0.076	3.70×10^{-6}	1.8 (1.4 to 2.3)
	Anti-CCP+	913	1901		0.112	0.076	1.80×10^{-5}	1.5 (1.3 to 1.9)
rs2476601	SE−	510	1901	A	0.126	0.093	2.40×10^{-3}	1.4 (1.1 to 1.7)
(*PTPN22*)	SE+	1376	1901		0.162	0.093	1.40×10^{-16}	1.9 (1.6 to 2.2)
	Anti-CCP−	402	1901		0.116	0.093	5.10×10^{-2}	1.3 (1.0 to 1.6)
	Anti-CCP+	913	1901		0.155	0.093	1.50×10^{-11}	1.8 (1.5 to 2.1)

Data for subphenotype analyses were available for 1886 Canadian cases.
*Versus healthy controls.
CCP, cyclic citrullinated peptide; NKAPL, NF-κB activating protein-like; RAF: risk allele frequency; SE, presence of shared epitope encoding *DRB1*0101, *0102,*0401,*0404, *0405,*0408, *0409, *1001* or **1402* alleles.

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 significant interactions 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 apparent between the RA-associated *PTPN22* rs2476601 variant and SE positivity (OR=1.34, 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* ($p=4.35\times10^{-10}$ and 2.51×10^{-11}) and the PTPN22 ($p=2.89\times10^{-15}$) 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,[10] 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 ($p<5.7\times10^{-8}$). Haplotype analysis reveals these six variants to all lie within one of these haplotype blocks in a region encompassing the NKAPL and three zing 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 ZNF193 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.

Table 5 HLA-DRB1 interaction with NKAPL or PTPN22 risk alleles

SNP	χ^2	p Value	OR (95% CI)
rs35656932	13.44	2.0×10^{-4}	0.67 (0.54 to 0.83)
rs13208096	11.27	8.0×10^{-4}	0.69 (0.55 to 0.86)
rs2476601	7.15	7.5×10^{-3}	1.34 (1.08 to 1.65)

Logistic regression analysis using case-only design was used to test for interactions between *HLA-DRB1* SE alleles and the *NKAPL* rs35656932 and rs13208096 SNPs or the *PTPN22* rs2476601 risk variant.
NKAPL, NF-κB activating protein-like.

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.[28] 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 effects 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 rs35656932, 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.[29] The gene encodes a 402 amino acid nuclear protein highly homologous to the *NKAP* that functions as a transcriptional repressor of NOTCH signalling in thymocytes and is required for haemopoietic stem cell maintenance and survival.[30 31] *NKAPL* is not functionally characterised, but shares with NKAP a domain critical to NKAP roles in transcriptional repression[30] 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, rs35656932, 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 and ZNF307 proteins encoded by the nearby genes. However, ZNF307 has transcriptional repressor activity and appears to target *NF-κB*, raising the intriguing, albeit highly speculative possibility that several genes at the *NKAPL* locus influence RA risk via effects on *NF-κB* signalling.[32 33]

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 emanate from highly plausible candidate genes likely to influence *NF-κB* 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 Genomic Medicine and a Tier 1 Canada Research Chair. PAG, CIA and YL were partially supported by US NIH Grant AR44422.

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 planned 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.

REFERENCES

1. **Firestein GS.** Evolving concepts of rheumatoid arthritis. *Nature* 2003;**423**:356–61.
2. **Plenge RM,** Seielstad M, Padyukov L, *et al*. TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. *N Engl J Med* 2007;**357**:1199–209.
3. **Remmers EF,** Plenge RM, Lee AT, *et al*. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med* 2007;**357**:977–86.
4. **Plenge RM,** Cotsapas C, Davies L, *et al*. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007;**39**:1477–82.
5. **Thomson W,** Barton A, Ke X, *et al*. Rheumatoid arthritis association at 6q23. *Nat Genet* 2007;**39**:1431–3.
6. **Barton A,** Thomson W, Ke X, *et al*. Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat Genet* 2008;**40**:1156–9.
7. **Raychaudhuri S,** Remmers EF, Lee AT, *et al*. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat Genet* 2008;**40**:1216–23.
8. **Raychaudhuri S,** Thomson BP, Remmers EF, *et al*. Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. *Nat Genet* 2009;**41**:1313–18.
9. **Gregersen PK,** Amos CI, Lee AT, *et al*. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat Genet* 2009;**41**:820–3.
10. **Stahl EA,** Raychaudhuri S, Remmers EF, *et al*. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010;**42**:508–14.
11. **Chen D,** Li Z, Yang Q, *et al*. Identification of a nuclear protein that promotes NF-kappaB activation. *Biochem Biophys Res Commun* 2003;**310**:720–4.
12. **Plenge RM.** Recent progress in rheumatoid arthritis genetics: one step towards improved patient care. *Curr Opin Rheumatol* 2009;**21**:262–71.
13. **Gregersen PK,** Olsson LM. Recent advances in the genetics of autoimmune disease. *Annu Rev Immunol* 2009;**27**:363–91.
14. **Begovich AB,** Carlton VE, Honigberg LA, *et al*. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004;**75**:330–7.
15. **Miretti MM,** Walsh EC, Ke X, *et al*. A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet* 2005;**76**:634–46.
16. **Gabriel SB**, Schaffner SF, Nguyen H, *et al*. The structure of haplotype blocks in the human genome. *Science* 2002;**296**:2225–9.
17. **Kazma R,** Babron MC, Génin E. Genetic association and gene-environment interaction: a new method for overcoming the lack of exposure information in controls. *Am J Epidemiol* 2011;**173**:225–35.
18. **Jawaheer D,** Seldin MF, Amos CI, *et al*. Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. *Arthritis Rheum.* 2003;**48**:906–16.
19. **Cornélis F,** Fauré S, Martinez M, *et al*. New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc Natl Acad Sci USA* 1998;**95**:10746–50.
20. **MacKay K,** Eyre S, Myerscough A, *et al*. Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum* 2002;**46**:632–9.
21. **Gregersen PK,** Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987;**30**:1205–13.
22. **Stenzel A,** Lu T, Koch WA, *et al*. Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. *Hum Genet* 2004;**114**:377–85.
23. **Newton JL,** Harney SM, Wordsworth BP, *et al*. A review of the MHC genetics of rheumatoid arthritis. *Genes Immun* 2004;**5**:151–7.
24. **Irigoyen P,** Lee AT, Wener MH, *et al*. Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles. *Arthritis Rheum* 2005;**52**:3813–8.
25. **Huizinga TW,** Amos CI, van der Helm-van Mil AH, *et al*. Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. *Arthritis Rheum* 2005;**52**:3433–8.
26. **Orozco G,** Pascual-Salcedo D, López-Nevot MA, *et al*. Auto-antibodies, HLA and PTPN22: susceptibility markers for rheumatoid arthritis. *Rheumatology (Oxford)* 2008;**47**:138–41.
27. **Mahdi H,** Fisher BA, Källberg H, *et al*. Specific interaction between genotype, smoking and autoimmunity to citrullinated alpha-enolase in the etiology of rheumatoid arthritis. *Nat Genet* 2009;**41**:1319–24.
28. **Lee HS,** Lee AT, Criswell LA, *et al*. Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus. *Mol Med* 2008;**14**:293–300.
29. **Su AI,** Wiltshire T, Batalov S, *et al*. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 2004;**101**:6062–7.
30. **Pajerowski AG,** Nguyen C, Aghajanian H, *et al*. NKAP is a transcriptional repressor of notch signaling and is required for T cell development. *Immunity* 2009;**30**:696–707.
31. **Pajerowski AG,** Shapiro MJ, Gwin K, *et al*. Adult hematopoietic stem cells require NKAP for maintenance and survival. *Blood* 2010;**116**:2684–93.
32. **Laity JH,** Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol* 2001;**11**:39–46.
33. **Li J,** Wang Y, Fan X, *et al*. ZNF307, a novel zinc finger gene suppresses p53 and p21 pathway. *Biochem Biophys Res Commun* 2007;**363**:895–900.

Supplementary methods

Subjects:

Among the cases included in this study, all fulfilled 1987 American College of Rheumatology diagnostic criteria for RA. Within the Canadian cohort, 75% of cases were female and data on autoantibody status, available only from Toronto patients, revealed 84% of cases to be positive for rheumatoid factor and 69% positive for anti-cyclic citrullinated peptide antibody. Within the US-based BRASS cohort were 863 anti-CCP positive RA cases and 1201 were matched healthy controls. Controls from all cohorts were healthy subjects with no history of autoimmune disease. Informed consent was obtained from all subjects using protocols approved by the local institutional review boards. Population stratification was not assessed in this study, but was found to be negligible in a previously reported GWAS (9) including a subset (598 cases/413 controls) of the Toronto cohort studied here. In that study, the median chi-square inflation value, lambda, across the genomewas 1.005 indicating essentially no deviation from the expected value of 1.0. In that study, we had removed regions of large inversions on chromsomes 8p (8.135-11.936 Mb) and the centromeric region of chromosome 17q21.31 (40-43 Mb) that are polymorphic in Caucasians and associated with increased LD. We also removed markers around the HLA region because they are associated with rheumatoid arthritis risk (from 24-36 Mb) and inclusion will inflate the median chi-square test value.

HLA-DRB1 testing

HLA-DRB1 alleles were determined using LAB Type HD Class II DRB1 Typing Test (Cat. no SSOH2B1: One Lambda Inc, Canoga Park, CA, USA), a DNA-based high-resolution HLA

typing method incorporating reverse sequence-specific oligonucleotide probes and Luminex technology. Briefly, the target DNA (*HLA-DRB1* gene) was amplified by PCR using group specific primers biotinylated for detection with streptavidin R-phycoerythrin conjugated probes. The PCR product was then denatured, hybridised to complementary flurophore-conjugated DNA probes and bound DNA detected on the Luminex system using HLA Fusion^{TM} Software Version 2.0 (One Lambda, Inc., Canoga Park, CA, USA) to map the reaction patterns to those associated with published *HLA* gene sequences and assign appropriate *HLA-DRB1* alleles. The shared epitope (SE) encoding DRB1 alleles tested were *DRB1*0101, *0102, *0401, *0404, *0405, *0408, *0409, *1001 and *1402.*

Genotyping

For genotyping on the Sequenom MassArray iPLEX platform, allele-extension products amplified from genomic DNA template were plated onto a SpectroCHIP and then subjected to mass spectrometry. Genotypings were called using SpectroCALLER software and genotype quality assessed using Sequenom Typer Analyzer. SNP annotation was based on NCBI dbSNP Build 36.3. Individual samples with genotype call rates <95% and SNPs with call rates <95%, minor allele frequencies <1%, or deviation from Hardy-Weinberg equilibrium (HWE) at $P < 0.0001$ were removed from the analysis. In total, 1368 cases and 1471 controls and 101 SNPs were included in the initial fine-mapping statistical analysis.

Imputation

HLA region imputation was conducted using data from a reference set of the 1958 UK Birth Cohort and CEPH samples comprising over 2500 individuals for whom both classical HLA genotypes and SNP genotypes are available. Imputation from the Canadian RA GWAS dataset (9) was conducted using the software package HLA*IMP (Dilthey, A. T., Moutsianas, L., Leslie, S., McVean, G. (2011): "HLA*IMP - An integrated framework for imputing classical HLA

alleles from SNP genotypes" Bioinformatics Advance Access, doi: 10.1093/bioinformatics/btr061.).

Supplementary Table S1. Association data from fine-mapping of the *NKAPL* locus

SNP	Position (bp)	Risk allele	RAF		OR	Confidence Level		PLINK *P*	Gene
			Case	Control		L95	U95		
rs17709145	28112535	G	0.07	0.05	1.41	1.12	1.78	3.84×10^{-3}	intergenic
rs175954	28119564	C	0.23	0.20	1.15	1.01	1.31	3.96×10^{-2}	intergenic
rs6913038	28152735	G	0.03	0.03	0.76	0.56	1.05	0.10	ZNF165
rs9393884	28154768	G	0.28	0.23	1.31	1.16	1.48	1.06×10^{-5}	ZNF165
rs17767294	28162177	G	0.08	0.08	1.01	0.83	1.23	0.96	ZNF165
rs10484404*	28163474	C	0.08	0.11	0.76	0.63	0.91	2.61×10^{-3}	ZNF165
rs203893	28170045	T	0.27	0.24	1.16	1.03	1.31	1.68×10^{-2}	ZNF165
rs169432	28186230	C	0.05	0.04	1.16	0.90	1.50	0.26	ZNF166
rs4713140	28205172	A	0.16	0.12	1.41	1.21	1.65	1.38×10^{-5}	ZNF435
rs1890809	28218457	T	0.08	0.11	0.71	0.59	0.85	2.28×10^{-4}	ZNF192
rs2622321	28225084	T	0.03	0.02	1.36	0.96	1.93	0.09	ZNF192
rs41269281	28242496	T	0.01	0.01	1.24	0.76	2.02	0.38	ZNF192
rs1225591	28256731	A	0.28	0.24	1.25	1.11	1.41	2.86×10^{-4}	intergenic
rs1225595	28259319	T	0.30	0.26	1.20	1.07	1.35	2.08×10^{-3}	intergenic
rs1150685	28264455	G	0.09	0.11	0.75	0.63	0.90	1.93×10^{-3}	intergenic
rs1225599	28273169	A	0.24	0.22	1.17	1.03	1.33	1.66×10^{-2}	intergenic
rs1150691	28276012	T	0.33	0.34	0.98	0.88	1.10	0.79	intergenic
rs13195291	28277220	A	0.12	0.07	1.61	1.33	1.94	5.03×10^{-7}	intergenic
rs1150694	28283132	T	0.38	0.38	1.00	0.90	1.12	0.94	intergenic
rs1150697	28283615	C	0.09	0.12	0.71	0.60	0.84	9.65×10^{-5}	intergenic
rs1233666	28284394	T	0.12	0.15	0.80	0.69	0.94	5.27×10^{-3}	intergenic
rs73400575	28292287	G	0.04	0.05	0.74	0.57	0.96	2.53×10^{-2}	ZNF193
rs1736894*	28294533	G	0.13	0.15	0.79	0.68	0.92	2.33×10^{-3}	ZNF193
rs72849214	28295194	A	0.06	0.05	1.12	0.89	1.41	0.34	ZNF193
rs73740587	28295698	A	0.12	0.14	0.85	0.73	1.00	0.05	ZNF193
rs9393908	28298809	A	0.15	0.11	1.40	1.20	1.64	2.18×10^{-5}	ZNF193
rs35656932	28299267	T	0.12	0.07	1.66	1.38	1.99	3.82×10^{-8}	ZNF193
rs73400581	28301446	G	0.04	0.05	0.72	0.56	0.93	1.22×10^{-2}	ZNF193
rs1150705	28301765	G	0.38	0.37	1.03	0.92	1.15	0.60	ZNF193

rs9461443	28302608	T	0.25	0.21	1.23	1.08	1.40	1.41×10^{-3}	ZNF193
rs12197427	28303500	A	0.04	0.04	0.86	0.66	1.13	0.28	ZNF193
rs34477097	28305165	T	0.25	0.22	1.17	1.04	1.33	1.23×10^{-2}	ZNF193
rs1150707	28305584	T	0.35	0.34	1.04	0.93	1.17	0.44	ZNF193
rs1233713	28306260	G	0.39	0.38	1.03	0.92	1.14	0.63	ZNF193
rs2299029	28306810	T	0.20	0.17	1.24	1.08	1.42	2.05×10^{-3}	ZNF193

Supplementary Table S1. Association data from fine-mapping of the *NKAPL* locus (cond)

			RAF			Confidence Level			
SNP	**Position (bp)**	**Risk allele**	**Case**	**Control**	**OR**	**L95**	**U95**	**PLINK *P***	**Gene**
rs11967137	28307743	G	0.19	0.17	1.18	1.03	1.35	1.98×10^{-2}	ZNF193
rs6917759	28309343	A	0.09	0.10	0.82	0.69	0.99	3.64×10^{-2}	ZNF193
rs13204012	28309510	A	0.12	0.07	1.63	1.35	1.95	1.53×10^{-7}	ZNF193
rs66487884	28310891	C	0.10	0.11	0.91	0.76	1.09	0.30	ZNF193
rs9380069	28311279	G	0.16	0.13	1.33	1.14	1.55	2.01×10^{-4}	ZNF193
rs11752073*	28312772	G	0.08	0.10	0.84	0.70	1.01	0.06	ZNF193
rs9393910	28316171	G	0.19	0.17	1.17	1.02	1.35	2.27×10^{-2}	ZNF307
rs1150711	28316514	T	0.32	0.32	0.99	0.89	1.11	0.90	ZNF307
rs7769054	28317176	C	0.12	0.13	0.89	0.75	1.04	0.14	ZNF307
rs7757215	28320468	C	0.05	0.05	0.88	0.69	1.13	0.32	ZNF307
rs13408	28320727	T	0.42	0.41	1.01	0.91	1.12	0.87	ZNF307
rs17720293	28322677	T	0.13	0.09	1.60	1.35	1.90	5.27×10^{-8}	ZNF307
rs1233710	28323425	C	0.06	0.07	0.92	0.75	1.14	0.45	ZNF307
rs2235359*	28323620	A	0.13	0.16	0.82	0.70	0.95	9.39×10^{-3}	ZNF307
rs68141011	28325776	T	0.16	0.13	1.25	1.08	1.46	3.09×10^{-3}	ZNF307
rs13200462	28326178	C	0.16	0.13	1.29	1.11	1.50	8.08×10^{-4}	ZNF307
rs1736904	28327249	T	0.21	0.19	1.13	0.99	1.29	0.08	ZNF307
rs1736895	28327805	T	0.41	0.41	1.02	0.91	1.13	0.75	ZNF307
rs72850297	28328557	T	0.07	0.06	1.21	0.98	1.50	0.08	ZNF307
rs10456362	28329795	A	0.16	0.13	1.30	1.12	1.51	5.29×10^{-4}	ZNF307
rs72850298	28330535	A	0.06	0.05	1.13	0.90	1.42	0.30	NKAPL
rs2185955	28330959	T	0.13	0.16	0.81	0.70	0.94	5.28×10^{-3}	NKAPL
rs7750106	28331659	C	0.04	0.05	0.79	0.61	1.01	0.06	NKAPL
rs72850299	28331704	C	0.06	0.05	1.09	0.87	1.37	0.46	NKAPL
rs12214383	28331710	T	0.41	0.40	1.02	0.92	1.13	0.73	NKAPL
rs9468321	28332504	A	0.04	0.04	0.95	0.72	1.25	0.71	NKAPL
rs13208096	28333290	G	0.12	0.07	1.61	1.34	1.94	3.69×10^{-7}	NKAPL
rs1531681	28334857	G	0.41	0.41	1.01	0.91	1.13	0.83	NKAPL
rs72850300	28334882	G	0.05	0.04	1.25	0.97	1.60	0.08	NKAPL
rs12000*	28335415	A	0.18	0.21	0.79	0.69	0.90	3.90×10^{-4}	NKAPL

rs3734564	28335839	T	0.04	0.04	1.13	0.86	1.48	0.39	NKAPL
rs1679709	28336321	T	0.16	0.13	1.32	1.14	1.53	2.72×10^{-4}	NKAPL
rs17720687	28336726	A	0.03	0.03	0.99	0.73	1.33	0.94	NKAPL
rs11751702	28337508	G	0.13	0.16	0.81	0.70	0.94	5.85×10^{-3}	NKAPL
rs1778508	28337860	G	0.16	0.13	1.28	1.10	1.49	1.15×10^{-3}	NKAPL
rs12215841	28340872	A	0.06	0.04	1.29	1.01	1.65	4.17×10^{-2}	NKAPL
rs2799077	28342576	T	0.16	0.13	1.28	1.10	1.49	1.62×10^{-3}	ZNF187

Supplementary Table S1. Association data from fine-mapping of the *NKAPL* locus (cond)

			RAF			Confidence Level			
SNP	**Position (bp)**	**Risk allele**	**Case**	**Control**	**OR**	**L95**	**U95**	**PLINK *P***	**Gene**
rs2799079	28343155	C	0.16	0.13	1.30	1.12	1.51	5.10×10^{-4}	ZNF187
rs73742509	28343447	G	0.04	0.05	0.81	0.63	1.04	0.10	ZNF187
rs67998226	28346038	C	0.12	0.08	1.66	1.39	1.99	2.48×10^{-8}	ZNF187
rs2179174*	28346948	T	0.17	0.21	0.79	0.69	0.91	7.76×10^{-4}	ZNF187
rs2394049	28347660	A	0.40	0.41	0.98	0.88	1.09	0.67	ZNF187
rs11965538	28347894	A	0.15	0.11	1.44	1.23	1.69	6.47×10^{-6}	ZNF187
rs1778483	28348970	C	0.31	0.31	1.00	0.89	1.13	0.96	ZNF187
rs72850302	28349377	G	0.06	0.05	1.12	0.89	1.41	0.33	ZNF187
rs1150726	28351021	C	0.34	0.34	0.98	0.87	1.09	0.69	ZNF187
rs72854503	28356041	G	0.07	0.06	1.22	0.98	1.50	0.07	ZNF187
rs1150724*	28358215	G	0.34	0.35	0.96	0.85	1.07	0.42	PGBD1
rs6901575	28358963	A	0.16	0.13	1.28	1.10	1.49	1.22×10^{-3}	PGBD1
rs11756111	28360838	T	0.14	0.17	0.81	0.69	0.93	4.28×10^{-3}	PGBD1
rs1150719	28364601	C	0.13	0.16	0.80	0.69	0.93	4.10×10^{-3}	PGBD1
rs9461448	28371700	T	0.04	0.05	0.82	0.63	1.07	0.14	PGBD1
rs1936365	28376431	G	0.16	0.14	1.15	0.99	1.34	0.06	PGBD1
rs1997660*	28377642	T	0.31	0.32	0.94	0.84	1.05	0.27	PGBD1
rs6456811	28378026	C	0.11	0.14	0.76	0.65	0.90	1.17×10^{-3}	PGBD1
rs853679	28404842	T	0.17	0.14	1.25	1.07	1.44	3.61×10^{-3}	ZNF323
rs16893970	28407274	T	0.15	0.18	0.79	0.68	0.92	1.86×10^{-3}	ZNF323
rs17312661	28408315	G	0.19	0.16	1.22	1.06	1.40	6.75×10^{-3}	ZNF323
rs7752448	28409078	G	0.12	0.09	1.29	1.08	1.53	4.63×10^{-3}	ZNF323
rs213243	28416765	T	0.04	0.05	0.87	0.67	1.12	0.27	ZNF323
rs11751928	28443357	C	0.14	0.16	0.80	0.69	0.93	4.47×10^{-3}	ZNF306
rs1052215	28456137	A	0.40	0.39	1.05	0.94	1.17	0.37	ZNF96
rs2859379	28466174	C	0.03	0.05	0.62	0.47	0.82	7.21×10^{-4}	ZNF96
rs2232422	28474241	C	0.08	0.09	0.86	0.70	1.04	0.12	ZNF96
rs6907950	28478225	T	0.31	0.29	1.11	0.99	1.25	0.09	ZNF96
rs17393342	28484002	A	0.08	0.07	1.31	1.07	1.61	9.88×10^{-3}	ZNF96

Legend: Association data are shown for all 101 SNPs tested in the fine-mapping of the *NKAPL* locus in 1368 RA cases and 1471 controls from the Toronto-based cohort. Tests for association were carried out

using *PLINK*. RAF = risk allele frequency. SNP annotation was based on NCBI dbSNP Build 36.3. The asterisk (*) indicates 8 SNPs across the NKAPL and surrounding locus that were also typed in our previous GWAS (9).

Supplementary Table S2. Conditional logistic regression analysis for six disease-associated SNPs at the *NKAPL* locus.

SNP	Position (bp)	Unconditioned P_{FDR} Value	*P* value conditioning on rs35656932	*P* value conditioning on rs35656932 and rs13208096
rs13195291	28277220	4.20×10^{-8}	0.530	0.213
rs35656932	28299267	2.20×10^{-9}	N/A	N/A
rs13204012	28309510	5.60×10^{-8}	**0.031**	0.055
rs17720293	28322677	5.10×10^{-8}	0.904	0.851
rs13208096	28333290	1.40×10^{-8}	**0.030**	N/A
rs67998226	28346038	2.20×10^{-9}	0.569	0.994

Conditional logistic regression analysis of the most strongly associated SNPs at the *NKAPL* locus was performed on the combined Toronto and Halifax cohort using SAS. The *P*-values for the associations using additive logistic regression conditional on the most significant rs 35656932 are shown. We then conditioned on both rs3656932 and fs13208096. N/A indicates the marker cannot be distinguished from the conditioned model.

Supplementary Table S3. Analysis of disease associations at the *HLA-DRB1*and *PTPN22* loci

SNP	Position (bp)	Chr.	Risk Allele	RAF		OR	Confidence Level		PLINK *P*	Gene
				Case	Control		L95	U95		
rs6910071	32390832	6	G	0.37	0.20	2.43	2.19	2.70	2.39×10^{-63}	HLA-DRB1tag
rs660895	32685358	6	G	0.39	0.19	2.65	2.38	2.94	2.04×10^{-75}	HLA-DRB1tag
rs2476601	114179091	1	A	0.15	0.09	1.75	1.52	2.02	2.72×10^{-15}	PTPN22

Legend: Association data are shown for the 3 indicated SNPs tested in the combined Toronto and Halifax cohort of 2078 cases and 1901 healthy controls. Tests for disease association were carried out using PLINK. RAF = risk allele frequency.

Supplementary Table S4. Logistic regression analysis of *NKAPL* locus SNPs conditioned on the *HLA DRB10401 tag SNPs rs660895 and rs6910071.**

SNP	Position (bp)	Gene	Unconditioned *P* value	*P* value conditional on rs660895	P value conditional on rs6910071
rs13195291	28277220	*ZNF193*	2.79×10^{-8}	4.49×10^{-12}	1.37×10^{-11}
rs35656932	28299667	*ZNF193*	7.34×10^{-10}	3.87×10^{-13}	1.19×10^{-12}
rs13204012	28309510	*ZNF193*	5.66×10^{-8}	9.53×10^{-12}	4.96×10^{-11}
rs17720293	28322677	*ZNF307*	4.29×10^{-8}	1.88×10^{-11}	4.95×10^{-11}
rs13208096	28333290	*NKAPL*	6.99×10^{-9}	3.08×10^{-12}	7.03×10^{-12}
rs67998226	28346038	*ZNF187*	6.53×10^{-10}	2.51×10^{-13}	3.18×10^{-13}
rs6910071	32390832	*HLA-DRB1 tag*	2.39×10^{-63}	N/A	N/A
rs660895	32685358	*HLA-DRB1 tag*	2.04×10^{-75}	N/A	N/A

Conditional logistic regression analysis wherein the six most strongly associated *NKAPL* SNPs were conditioned on each of the *HLA-DRB1*0401* tag SNPs was performed on the 2078 Canadian RA cases and 1901 controls. N/A indicates the marker cannot be distinguished from the conditioned model.

Supplementary Table S5. List of the imputed *HLA* alleles showing significant association with RA.

Imputed HLA alleles	*P-value*	OR (95% CI)
*HLA_DQA*0301*	3.58×10^{-25}	3.62 (2.84 - 4.62)
*HLA_DRB*0401*	5.09×10^{-21}	4.50 (3.29 - 6.16)
*HLA_DQB*0302*	2.58×10^{-11}	2.62 (1.98 - 3.48)
*HLA_DRB*1301*	2.59×10^{-7}	3.95 (2.34 - 6.67)
*HLA_DQB*0603*	3.15×10^{-7}	4.22 (2.43 - 7.32)
*HLA_DQB*0202*	3.83×10^{-7}	2.91 (1.93 - 4.39)
*HLA_DQA*0103*	6.06×10^{-7}	3.14 (2.00 - 4.92)
*HLA_DQA*0501*	6.96×10^{-7}	1.79 (1.42 - 2.25)
*HLA_DRB*0701*	1.94×10^{-6}	2.14 (1.56 - 2.93)
*HLA_DRB*0404*	2.83×10^{-6}	2.53 (1.71 - 3.72)
*HLA_C*0304*	3.13×10^{-6}	2.59 (1.74 - 3.87)
*HLA_DQA*0201*	4.00×10^{-6}	1.99 (1.49 - 2.66)
*HLA_C*0401*	5.90×10^{-5}	1.84 (1.37 - 2.48)

Legend: Association data are shown for the strongly associated *HLA* alleles ($p < 1 \times 10^{-4}$) emerging from imputation of 137 HLA alleles in 592 RA and 356 controls from the Toronto-based cohort. Tests for association were carried out using PLINK.

Supplementary Table S6A. Stepwise logistic regression in the seven SNP model.

SNP added to the model	Position (bp)	*P*-value	OR (95% CI)
rs35656932*	28299267	0.0004	2.33 (1.46-3.72)
rs762324	30563865	0.0001	3.09 (1.73-5.49)
rs3093662	31652168	0.006	2.24 (1.27-3.97)
rs2242653	31783744	0.004	1.81 (1.21-2.71)
rs2157337	32609122	<.0001	1.96 (1.40-2.74)
rs660895	32685358	<.0001	2.34 (1.64-3.33)
rs3130215	33182941	0.004	1.51 (1.14-1.99)

Legend: Stepwise logistic regression results based on analyses of 6 candidate SNPs from the fine mapping study (rs13195291, rs35656932, rs13204012, rs17720293, rs13208096, rs67998226), 379 strongly significant ($p < 1.0 \times 10^{-4}$) HLA SNPs from our prior RA genome-wide scan, and HLA alleles imputed in 592 RA and 356 controls from the Toronto-based cohort. * indicates the candidate SNP from the fine mapping study.

Supplementary Table S6B. Stepwise logistic regression with the *HLA-DRB1*0401* allele forced into the model.

SNP added to the model	Position (bp)	*P*-value	OR (95% CI)
*HLA-DRB1*0401*	-	0.04	1.73 (0.99-2.99)
rs35656932*	28299267	0.0002	2.42 (1.52-3.84)
rs12174433	30553455	0.0003	2.97 (1.65-5.35)
rs2242653	31783744	0.002	1.85 (1.24-2.75)
rs2157337	32609122	0.0001	1.91 (1.37-2.66)
rs660895	32685358	0.002	2.00 (1.28-3.13)

Legend: Stepwise logistic regression analyses of 6 candidate SNPs from the fine mapping study, 379 strongly significant HLA SNPs from our prior RA genome-wide scan, and the HLA imputed alleles with the *HLA-DRB1*0401* allele forced into the model. * indicates the candidate SNP from the fine mapping study.

Supplementary Table S7: Analysis *of HLA-DRB1-NKAPL and HLA-DRB1-PTPN22* interactions using a multinomial logistic model.

SNP	Case Group	Estimate	SE	1-df test P-value[a]	OR (95% CI)	2-df test P-value[b]
rs35356932-NKAPL	SE-	0.755	0.109	<.0001	2.13 (1.72 - 2.64)	
rs35356932-NKAPL	SE+	0.357	0.088	<.0001	1.43 (1.20 - 1.70)	2.51×10^{-11}
rs13208096-NKAPL	SE-	0.719	0.111	<.0001	2.05 (1.65 - 2.55)	
rs13208096-NKAPL	SE+	0.348	0.089	<.0001	1.42 (1.19 - 1.69)	4.35×10^{-10}
rs2476601-PTPN22	SE-	0.336	0.111	0.0026	1.40 (1.13 - 1.74)	
rs2476601-PTPN22	SE+	0.630	0.078	<.0001	1.88 (1.61 - 2.19)	2.89×10^{-15}

Legend: Multinomial logistic regression modeling was used to evaluate interaction effects. Risk outcomes were considered as low(0) in controls, intermediate (1) in SE-negative RA cases and high(2) in RA cases with 1or 2 SE alleles. SE: shared epitope. OR: odds ratio. [a]P-value from 1-degree of freedom tests; [b]P-value from 2-degree of freedom tests.

Supplementary Figure Legends:

Supplementary Figure S1
Study design and workflow.

Supplementary Figure S2. Diagram showing pairwise linkage disequilibrium of the *NKAPL* locus.
Linkage disequilibrium structure of the 372kb region encompassing the *NKAPL* gene on chromosome 6p22.1. The relationship of the genes across the region is shown at the top. The LD diagram below shows the haplotype block for the 56 genotyped SNPs with nominal significant association at $P < 0.05$, with intensity of shading proportional to the square of the correlation coefficient (r^2). The strongly associated SNPs rs35656932 and rs13208096 are highlighted in the red boxes.