

Allelic heterogeneity in *NCF2* associated with systemic lupus erythematosus (SLE) susceptibility across four ethnic populations

Xana Kim-Howard^{1,†}, Celi Sun^{1,†}, Julio E. Molineros¹, Amit K. Maiti¹, Hema Chandru¹, Adam Adler¹, Graham B. Wiley¹, Kenneth M. Kaufman², Leah Kottyan², Joel M. Guthridge¹, Astrid Rasmussen¹, Jennifer Kelly¹, Elena Sánchez¹, Prithvi Raj³, Quan-Zhen Li³, So-Young Bang⁴, Hye-Soon Lee⁴, Tae-Hwan Kim⁴, Young Mo Kang⁵, Chang-Hee Suh⁶, Won Tae Chung⁷, Yong-Beom Park⁸, Jung-Yoon Choe⁹, Seung Cheol Shim¹⁰, Shin-Seok Lee¹¹, Bok-Ghee Han¹², Nancy J. Olsen¹³, David R. Karp¹⁴, Kathy Moser¹, Bernardo A. Pons-Estel¹⁵, Edward K. Wakeland³, Judith A. James¹, John B. Harley², Sang-Cheol Bae⁴, Patrick M. Gaffney¹, Marta Alarcón-Riquelme¹, on behalf of GENLES[‡], Loren L. Looger¹⁶ and Swapan K. Nath^{1,*}

¹Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA, ²Cincinnati Children's Hospital Medical Center and the US Department of Veterans Affairs Medical Center, Cincinnati, OH, USA, ³Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA, ⁴Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea, ⁵Kyungpook National University Hospital, Daegu, Korea, ⁶Ajou University Hospital, Suwon, Korea, ⁷Dong-A University Hospital, Busan, Korea, ⁸Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Korea, ⁹Daegu Catholic University Hospital, Daegu, Korea, ¹⁰Daejeon Rheumatoid & Degenerative Arthritis Center, Chungnam National University Hospital, Daejeon, Korea, ¹¹Chonnam National University Hospital, Gwangju, Korea, ¹²Korea National Institute of Health, Osong, Korea, ¹³Division of Rheumatology, Department of Medicine, Penn State Medical School, PA, USA, ¹⁴Rheumatic Diseases Division, Department of Medicine, University of Texas Southwestern Medical Center, TX, USA, ¹⁵Sanatorio Parque, Rosario, Argentina and ¹⁶Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, VA, USA

Received July 19, 2013; Revised October 18, 2013; Accepted October 22, 2013

Recent reports have associated *NCF2*, encoding a core component of the multi-protein NADPH oxidase (NADPHO), with systemic lupus erythematosus (SLE) susceptibility in individuals of European ancestry. To identify ethnicity-specific and -robust variants within *NCF2*, we assessed 145 SNPs in and around the *NCF2* gene in 5325 cases and 21 866 controls of European-American (EA), African-American (AA), Hispanic (HS) and Korean (KR) ancestry. Subsequent imputation, conditional, haplotype and bioinformatic analyses identified seven potentially functional SLE-predisposing variants. Association with non-synonymous rs17849502, previously reported in EA, was detected in EA, HS and AA ($P_{EA} = 1.01 \times 10^{-54}$, $P_{HS} = 3.68 \times 10^{-10}$, $P_{AA} = 0.03$); synonymous rs17849501 was similarly significant. These SNPs were monomorphic in KR. Novel associations were detected with coding variants at rs35937854 in AA ($P_{AA} = 1.49 \times 10^{-9}$), and rs13306575 in HS and KR ($P_{HS} = 7.04 \times 10^{-7}$, $P_{KR} = 3.30 \times 10^{-3}$). In KR, a 3-SNP haplotype was significantly associated ($P = 4.20 \times 10^{-7}$), implying that SLE predisposing variants were tagged. Significant SNP–SNP interaction ($P = 0.02$) was detected between rs13306575 and rs17849502 in HS, and a dramatically increased risk (OR = 6.55) with a risk allele at each locus. Molecular modeling predicts that these non-synonymous mutations could disrupt NADPHO complex

*To whom correspondence should be addressed at: Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, 1025 N.E. 13th Street, Oklahoma City, OK 73104, USA. Tel: +1 4052717765; Fax: +1 4052714110; Email: swapan-nath@omrf.org

[†]These authors contributed equally to this work.

[‡]Members of GENLES are listed in the APPENDIX.

assembly. The risk allele of rs17849501, located in a conserved transcriptional regulatory region, increased reporter gene activity, suggesting *in vivo* enhancer function. Our results not only establish allelic heterogeneity within *NCF2* associated with SLE, but also emphasize the utility of multi-ethnic cohorts to identify predisposing variants explaining additional phenotypic variance ('missing heritability') of complex diseases like SLE.

INTRODUCTION

Systemic lupus erythematosus (SLE or lupus) is a clinically heterogeneous, systemic autoimmune disease affecting ~2 million people in the USA, about 90% of whom are women (1). The profound morbidity and mortality of SLE stem from chronic inflammation and multiple organ damage. Because lupus patients have weakened immune systems from standard immunosuppressant treatment (2–7) and immune dysregulation, infection is a leading cause of morbidity and mortality, accounting for >25% of SLE-related deaths (2,3,8–9).

Both innate immunity and global regulation of the immune system depend on generation of reactive oxygen species (ROS) such as superoxide ($O_2^{\bullet-}$) and hydrogen peroxide (H_2O_2). Pathogen clearance by phagocytosis depends on ROS production by phagocytes, i.e. neutrophils, monocytes and macrophages, and deficient ROS generation may give rise to infection susceptibility. Conversely, overproduction of ROS may lead to oxidative stress that stimulates autoimmune responses (10–12). Given the delicate ROS balance required for proper immune function, it is not surprising that mutations in ROS regulators may contribute to both immune dysfunction and susceptibility to chronic infections (13). At the core of the ROS production machinery is nicotinamide adenine dinucleotide phosphate oxidase (NADPHO), a multi-subunit membrane protein that catalyzes oxidation of NADPH. NADPHO consists of five core phagocytic oxidase ('phox') proteins, p22^{phox}/cytochrome b-245 alpha (*CYBA*), gp91^{phox}/Nox2/cytochrome b-245 beta (*CYBB*), p47^{phox} (*NCF1*), p67^{phox} (*NCF2*) and p40^{phox} (*NCF4*). The NADPHO complex is activated by one or more soluble GTPases in each phagocytic class (Supplementary Material, Fig. S1). The dominant NADPHO-activating GTPase is Rac1 in monocytes and Rac2 in neutrophils. Both GTPases appear to play a role in macrophages (14). A specific guanine nucleotide exchange factor (GEF), such as Vav1, is required to activate the GTPase (15).

Recently *NCF2*, which encodes p67^{phox}, was identified as an SLE susceptibility gene (16–18). An intronic SNP (rs10911363) was implicated in SLE pathogenicity in individuals with European ancestry (16,17). In Chinese individuals this SNP was not associated with SLE, but it was associated with arthritis and anti-nuclear antibody production (19). Fine-mapping by Jacob *et al.* (18) suggested that missense variant rs17849502 (H389Q) underlies *NCF2*-SLE association in European-derived populations. However, this could not be confirmed in other ethnic populations because of low minor allele frequency (MAF) and small sample sizes. Such disparate and inconclusive results are not uncommon for SLE, given the strong influence of ancestry on SLE predisposition and progression. SLE pathology is influenced by numerous molecular pathways and ethnicity-specific genetic associations (20,21), which suggests that distinct sets of genetic interactions modulate SLE risk in different populations. Also, prevalence and severity of SLE are markedly higher in individuals with African,

Asian or Hispanic ancestries compared with those with European ancestry (22–24). Thus, analyzing multiple ethnic populations is crucial for understanding the disease mechanisms of SLE (25).

To thoroughly investigate *SLE-NCF2* association in ethnically diverse populations, and to identify both ethnicity-specific and -robust causal variants, we analyzed a large multi-ethnic cohort. The objectives of this study were to (i) perform comprehensive analysis using dense fine-mapping, (ii) identify robust, independent SLE-predisposing variants and (iii) predict and validate the molecular effects of SLE risk alleles through bioinformatics, molecular modeling, and *in vitro* assays.

RESULTS

Study populations and genotypic data

Our large-scale association study included 12 256 participants from four ethnically diverse populations including European-American (EA), African-American (AA), Hispanic (HS) and Korean (KR) individuals (Supplementary Material, Table S1). All SLE patients satisfied American College of Rheumatology classification for SLE (26,27). Controls had no history of autoimmune disease and were considered healthy.

Our study design with workflow is shown in Figure 1. Individuals were genotyped on the Illumina ImmunoChip array (28) as a part of separate projects led by collaborators. Additional healthy controls were obtained from dbGaP (29). Subjects were excluded from analysis for low call rate (<98%), cryptic relatedness to another study participant (identity by descent >25%), or being outliers identified by principal component analysis (PCA) (Supplementary Material, Fig. S2). After quality control (QC) there were 4804 SLE cases and 7452 healthy controls: EA (1476 cases, 2693 controls), HS (829 cases, 1181 controls), AA (686 cases, 371 controls) and KR (1813 cases, 3207 controls) (Supplementary Material, Table S1).

A total of 145 SNPs in and around the *NCF2* region (Build 37; Chr1: 183,523,470–183,573,858) were examined. SNPs were removed if they were out of Hardy–Weinberg equilibrium ($P_{HWE} < 0.001$ in controls), had >10% missing genotype calls, clustered too poorly to separate genotype calls, or had MAF <0.5%. After QC, 118 SNPs were analyzed, including 74 SNPs in EA, 83 in HS, 94 in AA and 75 in KR. Case–control association studies were analyzed by χ^2 test, and odds ratios (ORs) and 95% confidence intervals (95% CIs) were calculated for each population using PLINK (30).

Initial case–control association analysis

The initial analysis showed significant association ($P < 0.05$) with SLE for 26 SNPs in EA, 5 in HS, 6 in AA, and 41 in KR (Supplementary Material, Table S2). The strongest association was at rs17849501 (A202A) in EA and HS [$P_{EA} = 5.23 \times 10^{-33}$, OR

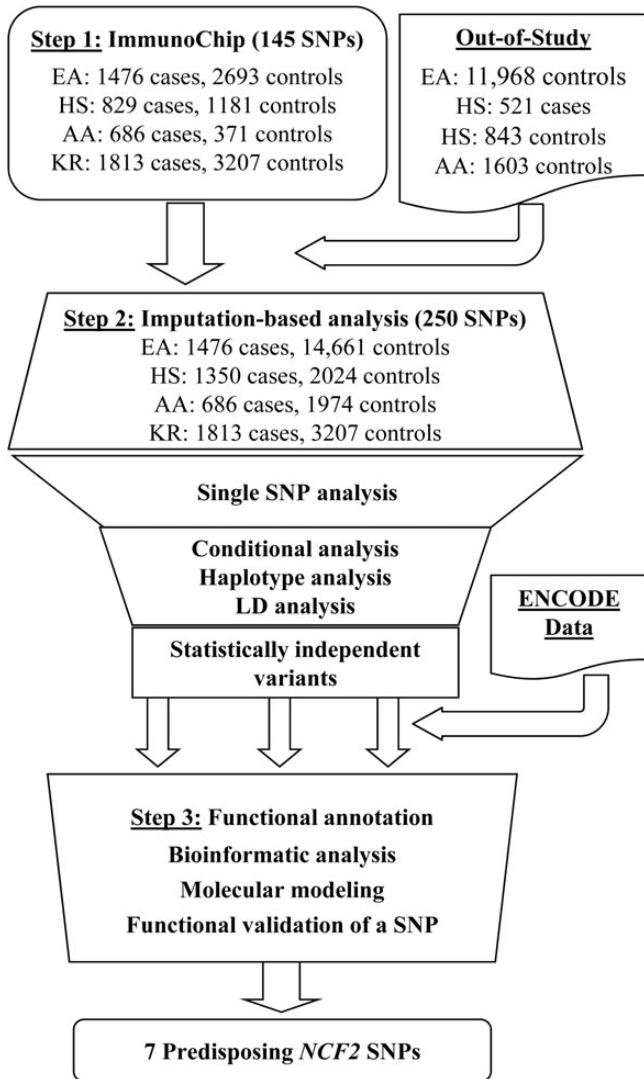


Figure 1. Workflow and study design of the present study.

(95% CI) = 2.73 (2.31–3.24); $P_{HS} = 2.97 \times 10^{-6}$, OR (95% CI) = 1.94 (1.46–2.56)]. Our results confirmed strong SLE association in EA and HS with the previously identified risk allele at rs17849502 (H389Q) (18) [$P_{EA} = 1.93 \times 10^{-28}$, OR (95% CI) = 2.48 (2.10–2.92); $P_{HS} = 7.01 \times 10^{-7}$, OR (95% CI) = 2.03 (1.53–2.69)]. Both of these SNPs were monomorphic in KR, and both failed to pass QC in AA due to low MAF (0.006 in controls). Novel association at non-synonymous SNP rs35937854 (V297A) was the strongest association in AA [$P_{AA} = 7.15 \times 10^{-4}$, OR (95% CI) = 2.20 (1.38–3.51)], but was monomorphic in other populations. In KR, the strongest association was at intronic SNP rs10911357 [$P_{KR} = 7.75 \times 10^{-4}$, OR (95% CI) = 1.16 (1.06–1.26)].

Imputation-based association analysis, adjusted for multiple testing and ancestry correction

To increase the statistical power of our analysis, we performed a comprehensive ethnicity-specific imputation-based analysis.

We incorporated an additional 11 968 EA and 1603 AA healthy out-of-study controls from dbGaP, as well as 521 HS SLE cases and 843 healthy HS controls obtained through collaboration (Supplementary Material, Table S1; Fig. 1). We also increased fine-mapping density by imputing additional *NCF2* SNPs from 1000 Genomes Project data using MACH-ADMIX (31). QC criteria described earlier were applied to these samples and SNPs. Additionally, SNPs had to pass imputation quality metric ($R_{sq} > 0.7$).

After imputation and QC, 250 SNPs were available for analysis (Supplementary Material, Table S3). We adjusted the significance threshold of $\alpha = 0.05$ to account for multiple testing. Spectral decomposition of the pair-wise LD matrices (32,33) was used to account for inter-SNP correlation due to linkage disequilibrium (LD), allowing us to determine the effective number of independent SNPs separately for each population. Based on this calculation, we determined a population-specific adjusted experiment-wise significance threshold at the 5% level: EA ($P \leq 1.33 \times 10^{-3}$, 152 total SNPs and 6 independent SNPs), HS ($P \leq 1.70 \times 10^{-3}$, 134, 11), AA ($P \leq 4.95 \times 10^{-4}$, 214, 5) and KR ($P \leq 2.00 \times 10^{-3}$, 164, 5).

Imputation-based analysis supported our initial results. rs17849502 was the most strongly associated SNP in EA [$P_{EA} = 1.01 \times 10^{-54}$, OR (95% CI) = 2.57 (2.28–2.91)] and was also highly significant in HS [$P_{HS} = 3.68 \times 10^{-10}$, OR (95% CI) = 2.02 (1.62–2.53)]. In both populations, rs17849501 was similarly associated [$P_{EA} = 4.45 \times 10^{-54}$, OR (95% CI) = 2.56 (2.26–2.90); $P_{HS} = 3.34 \times 10^{-10}$, OR (95% CI) = 2.02 (1.61–2.52)], and was in strong LD with rs17849502 (Supplementary Material, Fig. S3). Since the MAF in AA was too low for reliable imputation (MAF_{rs17849502} = 0.0066, MAF_{rs17849501} = 0.0056 in controls), we genotyped rs17849502 and rs17849501 in a subset of 603 AA cases and 384 AA controls using a TaqMan assay. Despite the reduced sample size and low MAF, both SNPs were significantly associated with SLE [$P_{rs17849502} = 0.02$, OR (95% CI) = 2.76 (1.14–6.68); $P_{rs17849501} = 0.007$, OR (95% CI) = 3.88 (1.36–11.10)]. Consistent with our initial results, the strongest association in AA was at rs35937854 [V297A; $P_{AA} = 1.49 \times 10^{-9}$, OR (95% CI) = 2.34 (1.76–3.11)]. In KR, the strongest SLE association was an imputed intronic SNP, rs10911359 [$P_{KR} = 6.82 \times 10^{-4}$, OR (95% CI) = 1.16 (1.06–1.26)].

Robustness of our analyses was assessed using covariate adjusted logistic regression implemented in PLINK (see Methods and Materials). Using ADMIXTURE (34) software, we detected three ancestral components as optimal for describing AA, KR and EA, whereas four ancestral components best described HS. Individual ancestry proportions were used as covariates to correct for potential effects of population structure on association (Supplementary Material, Table S3). Results were confirmed by principal component analysis (PCA, implemented in GCTA) (35) using the first three principal components covariates (Supplementary Material, Table S3). For both covariate adjusted result sets the magnitude of significance was consistent with the unadjusted results, with the exception of rs17849502/rs17849501 in EA, and rs13306575 in HS, where ancestry might have a small impact on the strength of association (Supplementary Material, Table S3). Both associations remain significant following ancestry adjustment.

Identification of independent variants in four ethnicities

We performed conditional analysis to detect independently associated variants in each population (Supplementary Material, Table S4). Population-specific haplotypes were constructed using independent SNPs identified through conditional analysis. Interactions between independently associated SNPs in each population were also assessed using additional methods (described below).

SLE association of the *NCF2* region in EA was explained by rs17849502 (H389Q) and rs17849501 (A202A). Conditioning on either SNP explained SLE significance in EA (using threshold $P > 1.33 \times 10^{-3}$) (Supplementary Material, Fig. S4). Because of the strong LD between these SNPs ($D' = 1.0$, $r^2 = 0.99$, 9743 bp apart), we could not dissect separate contributions of each SNP (Supplementary Material, Fig. S3) and considered these SNPs together.

In HS both rs17849502 and rs17849501 were also highly significant [$P_{HS} = 7.04 \times 10^{-7}$, OR (95% CI) = 1.56 (1.31–1.86)] and in high LD ($D' = 1.0$, $r^2 = 0.98$). Additionally, non-synonymous variant rs13306575 (R395W) was independently associated with SLE (Table 1) [$P_{HS} = 7.04 \times 10^{-7}$, OR (95% CI) = 1.56 (1.31–1.86)]. rs13306575 was also significant in KR [$P_{KR} = 3.30 \times 10^{-3}$, OR (95% CI) = 1.43 (1.13–1.83)] in the same direction (Fisher’s combined P -value for $P_{HS\&KR} = 4.85 \times 10^{-8}$). The omnibus haplotype association was highly significant ($P_{hap} = 4.75 \times 10^{-14}$) (Supplementary Material, Table S5). Conditioning on haplotypes with a risk allele at only one SNP did not explain the omnibus haplotype significance ($P_{cond:GT} = 2.33 \times 10^{-9}$; $P_{cond:AG} = 5.60 \times 10^{-10}$), but conditioning on the protective haplotype ‘GG’ accounted for the entire omnibus haplotype association ($P_{cond:GG} = 0.09$) (Supplementary Material, Fig. S5). Haplotype analysis including rs17849501/rs13306575 produced similar results (data not shown). Thus in HS, SLE association is explained by rs13306575 and rs17849502/rs17849501, the latter acting as an indistinguishable effect.

We observed a significant SNP–SNP interaction between independent, missense SNPs rs17849502 and rs13306575 using a logistic regression model implemented in R (PC-corrected $P = 0.024$, ancestry-corrected $P = 0.025$). This result was confirmed by parametric FITF (36), epistasis test ($P = 0.018$), PLINK two-locus test ($P = 0.02$) and GATA ($P = 0.038$). We confirmed this interaction (sign test $P = 1 \times 10^{-3}$) using non-parametric generalized multifactor-dimensionality reduction (GMDR). In GMDR, the best-fitted interaction model was between these two SNPs, with a cross-validation consistency of 9 out of 10, and a testing accuracy of 55%. We then compared cases and controls with no risk alleles to individuals with a risk allele at both SNPs and found a dramatic increase in SLE risk [OR (95% CI) = 6.55 (2.84–15.11)]. Possessing a risk allele at only one SNP showed a marginal increase in SLE risk [risk at rs13306575, OR = 1.52 (1.25–1.85); risk at rs17849502, OR = 1.78 (1.38–2.28)].

In AA, non-synonymous rs35937854 (V297A) and intronic rs34680162 were independently associated with SLE (Table 1). Conditional analysis on overall haplotype association ($P_{hap} = 2.93 \times 10^{-11}$) confirmed that neither SNP fully explained association ($P_{cond:rs35937854} = 1.15 \times 10^{-4}$; $P_{cond:rs34680162} = 2.20 \times 10^{-9}$) (Supplementary Material, Fig. S6 and Table S5). No observed haplotypes contained risk alleles at both SNPs.

Table 1. Single SNP allelic association results for four populations. Seven SNPs were independent in at least one ethnicity (bold). Three SNPs caused a change of amino acid (non-synonymous), one was synonymous, and the remaining two were intronic. Minor allele frequencies for SLE cases (MAF-A) and healthy controls (MAF-U), along with P -values and odds ratios (OR) are presented.

SNP (function)	Position (feature)	Allele	European American (1476/14,661)*		Hispanic (1350/2024)*		African American (686/1974)*		Korean (1813/3207)*	
			MAF-A	P-value	MAF-U	P-value	MAF-A	P-value	MAF-U	P-value
rs13306575 (R395W)	183532437 exon 14	A/G			0.066	7.04x10⁻⁷	monomorphic		0.034	3.30x10⁻³
rs17849502 (H389Q)	183532580 exon 13	T/G	0.120	1.01x10⁻⁵⁴	0.035	3.68x10⁻¹⁰	0.008	0.02 ⁺		monomorphic
rs35937854 (A297V)	183536089 exon 10	G/A				monomorphic	0.029	1.49x10⁻⁹	0.066	2.34
rs34680162	183538434 intron 8	C/A				monomorphic	0.025	2.33x10⁻⁴	0.045	1.83
rs17849501 (A202A)	183542323 exon 6	T/C	0.119	4.45x10⁻⁵⁴	0.036	3.34x10 ⁻¹⁰	0.005	0.007 ⁺	0.034	3.88
rs10911359	183543053 intron 5	C/T	0.513	0.256	0.375	0.825	0.649	0.417	0.95	0.388
rs34423782	183554714 intron 4	C/T	0.936	0.185	Failed QC		0.824	0.246	1.10	0.986
										1.43
										1.16
										1.73x10⁻³

*Sample sizes (cases/controls); ⁺ Due to low MAF, P -value was determined in a subset of N cases and N controls sequenced by Taqman, and P -value met the threshold for significance in this experiment; OR = odds ratio

Conditioning on haplotypes including risk alleles at only one SNP did not explain the omnibus haplotype significance ($P_{\text{cond:AC}} = 2.20 \times 10^{-9}$; $P_{\text{cond:GA}} = 1.15 \times 10^{-4}$), but conditioning on the protective haplotype 'AA' explained haplotypic association ($P = 0.28$). No SNP-SNP interactions were observed.

In KR, we identified three novel independently associated *NCF2* SNPs: non-synonymous variant rs13306575, and intronic variants rs10911359 and rs34423782 (Table 1) in low LD with one another (Supplementary Material, Fig. S3). No SNP fully explained the overall omnibus haplotype association ($P_{\text{hap}} = 4.20 \times 10^{-7}$; $P_{\text{cond:rs13306575}} = 9.50 \times 10^{-6}$; $P_{\text{cond:rs10911359}} = 3.14 \times 10^{-4}$; $P_{\text{cond:rs34423782}} = 2.72 \times 10^{-5}$). Likewise, conditioning on single haplotypes did not explain overall haplotype significance ($P_{\text{cond:GTT}} = 2.72 \times 10^{-5}$; $P_{\text{cond:GCC}} = 3.14 \times 10^{-4}$; $P_{\text{cond:ATC}} = 1.26 \times 10^{-5}$; $P_{\text{cond:GTC}} = 3.17 \times 10^{-5}$) (Supplementary Material, Fig. S7). No haplotypes containing more than one risk allele were observed, and no SNP-SNP interactions were observed. Both rs10911359 and rs34423782 lie within active chromatin regions in GM12878 cells (a lymphoblastoid cell line); both risk alleles disrupt CpG dinucleotides within tracks of methylated sequence (ENCODE/MeDIP track).

We assessed the heterogeneity of OR across populations for independently associated SNPs that passed QC in at least two populations. Control allele frequencies were consistent with HapMap frequencies for each population. We found that for all but one SNP the ORs were homogeneous across populations. For rs10911359, which was only significant in Koreans, the ORs were heterogeneous ($P = 0.0056$).

In silico functional analysis

Together, the conditional analyses identified seven associated SNPs with six independent effects (counting rs17849501 and rs17849502 together) in the *NCF2* region; three of these were non-synonymous variants. Because each non-synonymous mutation was monomorphic in at least one population, a single amino acid change (e.g. H389Q) (18) cannot explain the *NCF2* association in all populations. To better understand how genetic variation in these populations influences SLE susceptibility, independently associated variants, coding and non-coding, were scanned *in silico* to assess their possible functional effects on *NCF2* or directly adjacent genes.

We focused on probes targeting *NCF2* and its adjacent genes, *SMG7* and *ARPC5*, using genotype data from HapMap and the 1000 Genomes Project and expression data from GeneVar (37,38) for expression quantitative trait locus (eQTL) analysis. A linear model was fitted for population-specific eQTL expression data versus *NCF2* SNP genotypes for 153 HapMap samples using PLINK. Five SNPs (rs13306575, rs17849502, rs34680162, rs17849501, rs10911359; Supplementary Material, Fig. S8, Table S6) were significantly associated with *cis*-eQTLs at the *SMG7* probe (GI_42476073-A). These results were also assessed using data from the mRNA browser (39,40). Since none of these SNPs were available in the mRNA browser, we used tagged SNPs that were in LD (r^2 or $D' \geq 0.8$) with our SNPs (Supplementary Material, Table S6) to assess associations with mRNA transcript abundance.

Four tagged SNPs, rs1044879 ($D' = 1$, $r^2 = 1$ with rs17849501 and rs17849502 in CEU, $P_{\text{mRNA}} = 1.4 \times 10^{-17}$), rs796860 ($D' = 1$, $r^2 = 0.05$ with rs10911359 on YRI,

$P_{\text{mRNA}} = 5.0 \times 10^{-4}$), rs2296164 ($D' = 1$, $r^2 = 1$ with rs10911359 in CEU and CHB + JPT, $P_{\text{mRNA}} = 2.6 \times 10^{-10}$) and rs2274064 ($D' = 1$, $r^2 = 0.02$ with rs10911359 and rs34423782 in CHB and JPT, $P_{\text{mRNA}} = 6.2 \times 10^{-10}$) were significantly associated with transcript abundance of *SMG7*, but not *NCF2* or *ARPC5*. Of note, although two tag-SNP pairs (rs796860- rs10911359 and rs2274064- rs34423782) had low correlation (r^2), their minor alleles were on the same haplotype ($D' = 1$). rs10911359 is in perfect correlation with rs2296164, leaving only rs34423782 imperfectly tagged. These significant associations between tagged SNPs in *NCF2* and *SMG7* transcript abundance demonstrate that genetic variants within *NCF2* associated with SLE may have functional effects on expression of the adjacent gene *SMG7*. *SMG7* encodes a component of the mRNA 'decay body', which mediates nonsense-mediated mRNA decay (NMRD) (41). Disruptions to NMRD have been implicated in auto-antibody development, in particular, to lupus auto-antigen La (42). Furthermore, *SMG7* levels have been shown to correlate negatively with lupus anti-nuclear antibody titers (43), suggesting that *SMG7* in particular and NMRD in general may play a protective role against auto-antigen generation and SLE progression.

Molecular modeling of the novel coding variants

We modeled protein structural effects of non-synonymous rs13306575 (R395W), which was independently associated with SLE in HS and KR. The risk allele at rs13306575 results in an R395W mutation in exon-14 of the p67^{phox} protein encoded by *NCF2*. A previous crystal structure of the PB1 domain complex of p40^{phox} and p67^{phox} (44) showed that R395 forms a salt bridge with the carbonyl oxygen of P339 in p40^{phox} and R395W may disrupt this salt bridge (Fig. 2A). Additionally, the tryptophan side-chain introduces significant clashes within the PB1 domain, likely destabilizing the domain and its interactions.

Similarly, the risk allele at rs35937854 produces a V297A substitution in exon-10 and is predicted to disrupt a highly conserved binding site (ENCODE/HMR track) for the transcriptional activator Myb (Myeloblastosis oncogene). Myb is critical for hematopoiesis, particularly monocyte-macrophage differentiation and proliferation (45). The V297A substitution in p67^{phox} is located in the first SH3 domain, which interacts with gp91^{phox} (46). This mutation is predicted to destabilize SH3 domain folding (Fig. 2B), which may in turn weaken the interaction with gp91^{phox}. These observations suggest that rs13306575 and rs35937854 risk alleles may have both direct and indirect effects on protein function.

Luciferase reporter assay with the synonymous variant

Bioinformatic analysis demonstrated that the synonymous variant rs17849501 (A202A) may act as a *cis*-regulatory element to influence protein and mRNA expression. During protein translation, the wild-type codon (GCG) pairs with the least abundant alanine tRNA, whereas the complementary tRNA for the mutant codon (GCA) is abundant (www.kazusa.or.jp/codon), suggesting that the risk allele could increase translation efficiency. Moreover, rs17849501 is in a highly conserved region annotated as active chromatin in GM12878 cells

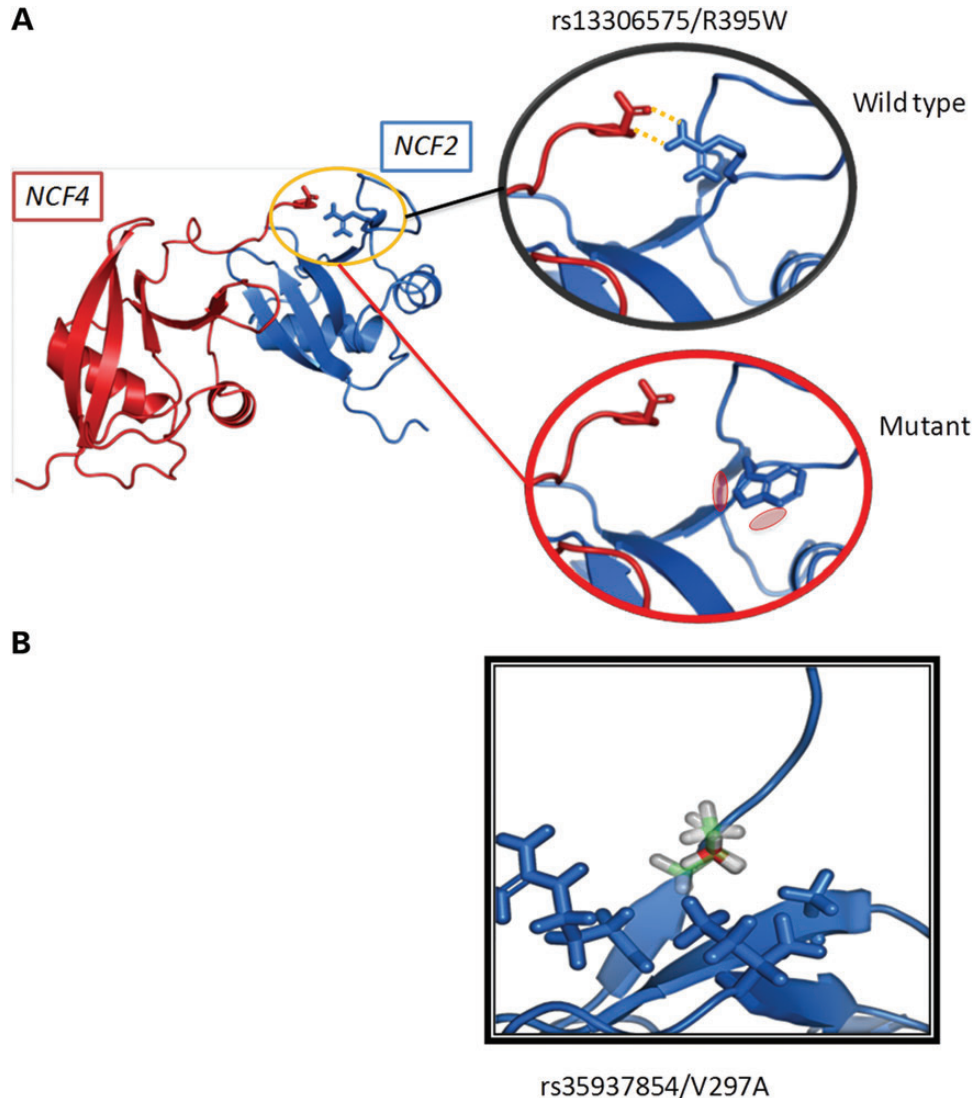


Figure 2. (A) Effect of rs13306575 (R395W) on the interactions between p67^{phox} and p40^{phox}. Structure (PBD# 1OEY) of p67^{phox}-R395 (wild-type) (top) making an ion pair with the carbonyl oxygen of P339 in p40^{phox}; lost in R395W (mutant); the tryptophan side also introduces clashes (bottom). (B) Effect of rs35937854 (V297A) on the conformation and stability of the first SH3 domain.

(ENCODE/Broad Histone track). This region is notably enriched for trimethylated histone H3-Lys36 (H3K36Me3), a transcriptional elongation signal, and trimethylated histone H3-Lys27 (H3K27Me3), frequently associated with transcriptional silencing. Furthermore, the cytosine in the rs17849501 wild-type allele is in a methylated CpG motif (ENCODE/MeDIP-Seq track). Based on this evidence, we directly assessed transcriptional activity of the rs17849501 locus with a luciferase expression assay. The wild-type rs17849501 locus enhanced transcription in MonoMac-6 cells by ~15-fold over background. The rs17849501 locus carrying the risk allele boosted expression by a further ~40% ($P = 9.7 \times 10^{-8}$) (Fig. 3). Thus the rs17849501 risk allele may increase transcriptional enhancer activities by suppressing silencing function of the non-risk allele, perhaps through epigenetic interactions. In combination with potential increased translational efficacy, rs17849501 could dysregulate NADPHO even though the mutation does not change protein sequence.

DISCUSSION

In the present study, we demonstrated that multiple independent and robustly associated variants within the *NCF2* gene, both coding and non-coding, are predicted to influence SLE susceptibility. Intriguingly, the set of SLE-associated SNPs was quite different in each population.

This work also offers insight into possible mechanisms by which SLE-associated variants in *NCF2* may influence SLE risk. First, associated SNPs have the potential to influence epigenetic regulation of this locus. For example, the synonymous rs17849501 variant seems to disrupt an endogenous silencer module. This SNP is in strong LD with rs17849502 and may contribute to the latter's SLE association. Secondly, several of the SNPs identified here may control expression of the adjacent *SMG7* gene, which has also been implicated in SLE risk. Thirdly, three of the independently associated variants are non-synonymous mutations with the potential to disrupt protein

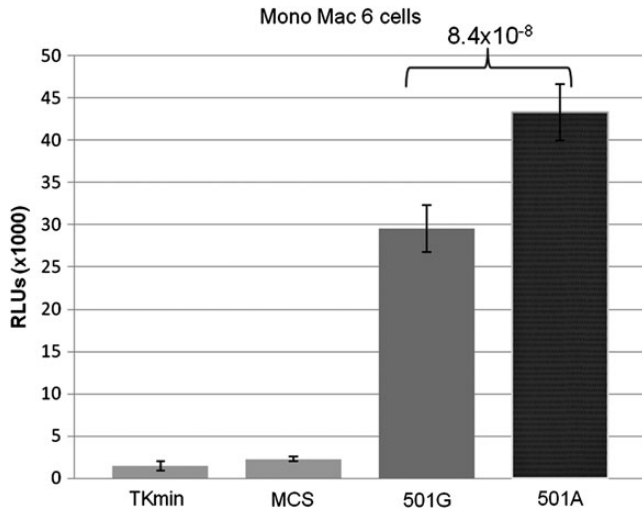


Figure 3. Enhancer activity of exonic DNA sequences carrying the rs17849501 wild-type (501G) or risk allele (501A). The 106-mer DNA sequences carrying rs17849501 non-risk (G) and risk (A) alleles were cloned into an enhancer/silencer assay vector (TKmin-MGL.1, Xactagen, USA) for luciferase activity carrying a minimal *TKmin* promoter. These clones were transiently transfected by electroporation to MonoMac-6 cells, and luciferase activity was measured at 6 h after transfection (Klan and Steinhilber, 2003). Transfection efficiency was more than >90% in all cases (measure of *CMV* promoter control, data not shown). TK_{min} = vector only, MCS = vector with multiple cloning sites, 501G = vector with wild-type allele, 501A = vector with risk allele. Bars represent the average promoter activity and standard deviation for nine replicates, expressed in relative luminescence units.

function. For example, the novel amino acid mutation caused by rs35937854 (V297A) likely destabilizes the first p67^{phox} SH3 domain and presumably its binding interactions. Finally, in HS two of the non-synonymous SNPs were polymorphic. Since multiple interacting loci may contribute to SLE susceptibility, we assessed SNP–SNP interaction (epistasis) to better understand *NCF2* association with SLE. The SNPs that showed the strongest interaction, rs17849502 and rs13306575, yield amino acid substitutions that are separated by just 6 positions in the PB1 domain of p67^{phox}. The rs17849502 mutation has been shown to adversely affect binding between the p67^{phox}-PB1 domain and Vav1 (47), and the rs13306575 mutation likely affects binding of the p67^{phox}-PB1 domain to p40^{phox} (44). Genetic interaction between these variants suggests that destabilizing both of these interactions may have a synergistic disruptive effect on the formation and overall activity of the NADPHO complex.

Functional mutations in NADPHO genes that adversely affect complex assembly have been shown to disrupt ROS production and predispose patients to primary immunodeficiency, such as chronic granulomatous disease (CGD) (48). Assembly of the NADPHO complex requires precise coordination of several protein subunits (49–51), and multiple mutations affecting the expression and/or function of the various components of NADPHO (*CYBA*, *CYBB*, *NCF1*, *NCF2*, *NCF4*) (52–62) are strongly predisposing for CGD. In addition, rs13306575 (R395W), an SLE-associated SNP identified here, has been previously associated with inflammatory CGD (63,64), although the rs13306575 mutation was paired with an in-frame deletion of three amino acids in the N-terminal tetratricopeptide repeat domain, which binds Rac-GTP (65). One CGD patient with the R395W mutation had no detectable p67^{phox} in her neutrophils

based on western blot (64). Several other inflammatory diseases have also been associated with NADPHO genes, including inflammatory bowel disease with *NCF2* (66), Crohn's disease with *NCF4* (67,68) and rheumatoid arthritis with *NCF2*, *NCF4* and *RAC2* (69). A recent large-scale resequencing study demonstrated that rs17849502/rs17849501 is also associated with celiac disease (70). This indicates that *NCF2* may be a common autoimmunity gene, increasing susceptibility for multiple autoimmune/inflammatory phenotypes (70). Because NADPHO is a central regulator of ROS levels, which are critical for regulation of inflammation, adaptive immunity, intracellular signaling, chemoattraction and autophagy (71–77), it is not surprising that mutations affecting this complex are implicated in inflammatory diseases.

Interestingly, recent reports demonstrated that physiologically appropriate NADPHO function is crucial for the formation of neutrophil extracellular traps (NETs) (78–80) and have linked improper clearance of NET debris with auto-antigen production in SLE (80–82). NETs are extracellular structures composed of chromatin and granular proteins that are secreted by neutrophils to entrap and kill microorganisms (83,84). Monocytes from SLE patients have a reduced ability to degrade *in vitro* generated NETs, and the abundance of NETs in SLE patients is correlated with serum levels of immunogenic complexes (82). Furthermore, patients that failed to degrade NETs had more active disease (85). Indeed, Campbell *et al.* (80) found that Nox2-deficient (and thus NADPHO-deficient) mice had severe lupus, including increased renal disease and elevated autoantibody profiles. These results suggest that failure to undergo normal NADPHO-dependent cell death may increase immunogenic cellular debris, breaking self-tolerance. The *NCF2* variants discovered here could act through a similar mechanism, by disrupting ROS production, leading to accumulated NET debris and auto-antigenicity.

Our results suggest that *NCF2* influences SLE risk through a complex balance between many underlying predisposing variants, some of which perturb expression and some of which affect the structure and function of the translated protein. Interactions with other genes would be expected to alter this balance, causing SNPs to have discordant effects on SLE risk in different populations when their intergenic interactions varied between populations. The potentially functional SNPs identified here have dramatically different effects on SLE risk in different populations, underscoring the importance of using large, multi-ethnic studies not only in elucidating the underlying mechanisms of SLE, but also in identifying predisposing variants explaining additional phenotypic variance (i.e. missing heritability) of complex diseases such as SLE. Our bioinformatics and functional work strongly suggests that these SNPs contribute directly to disruption of NADPHO. However, it is possible that the observed allelic heterogeneity results, at least in part, from unidentified causal SNPs tagged by the SNPs associated in this study. Further experiments may clarify this issue.

In summary, our strategy using a large, multi-ethnic case–control SLE cohort for dense ImmunoChip-based fine-mapping with imputation-based association analysis, followed by conditional analysis, bioinformatics analysis and molecular modeling, identified 6–7 (two behave dependently) potentially functional variants in *NCF2*. For many of the variants, we have implicated mechanisms by which they may impact NADPHO function.

However, no variant was associated in all four populations, and some variants showed significant associations only within a single population. This high degree of allelic heterogeneity in *NCF2* association with SLE suggests that several risk variants may be undiscovered. In particular, given that rare variants could not be successfully imputed, this study provides both the groundwork and the impetus for full gene sequencing in multiple ethnically diverse populations in a follow-up case–control study. Thus, understanding how these SNPs contribute to SLE can focus future functional experiments vital to uncovering how genetic variability predisposes for disease pathogenesis.

MATERIALS AND METHODS

Study samples, genotyping and quality control

The study samples were recruited through the Lupus Family Registry and Repository (86) at the Oklahoma Medical Research Foundation (OMRF), the Cincinnati Children’s Hospital Medical Center (CCHMC), the University of Texas SouthWestern (UTSW), the Center for Genomics and Ontological Research, Spain (GENYO) and the Hanyang University, South Korea (HU). Protocols for this study were approved by the Institutional Review Boards (IRBs) where participants were recruited (OMRF, UTSW, CCHMC, GENYO and HU). Every individual, including SLE cases and healthy controls, provided written consent at enrollment. All SLE patients met a minimum of 4 out of 11 1997 American College of Rheumatology revised classification criteria for SLE (27,87). Controls were healthy individuals, who did not have lupus or any other autoimmune disease. Most study controls were recruited through the Lupus Family Registry and Repository (LFRR) (86). All controls were recruited from the same countries and areas where patients were recruited. Whenever possible, they were selected to have the same self-reported ethnicity as the patient population. Controls were recruited with IRB approval by participating physicians, or were recruited from health fairs, community health clinics, or area schools. Out of study GWAS controls were obtained with permission through dbGaP.

SLE cases and healthy controls were genotyped on the Illumina ImmunoChip array (28) (Supplementary Material, Table S1). For quality control (QC), subjects were excluded from analysis if they had >10% missing genotyping, were related to any other study participant (>25% identical by descent), or were population stratification outliers based on PCA. SNPs were removed for >10% missing genotyping, being out of Hardy–Weinberg equilibrium ($P_{\text{HWE}} < 0.001$ in controls), or for poor clustering.

We increased statistical power to detect lupus-associated variants by incorporating out-of-study healthy controls obtained through dbGaP (Supplementary Material, Table S1). For AA, we added controls from the Health and Retirement Study (HRS) genotyped on the Illumina OMNI 2.5 platform. For EA, we incorporated control genotypes from the GENIE UK-ROI Diabetic Nephropathy GWAS (GoKIND), Genes and Blood Clotting Study (GABC) and NGRC Parkinson’s Disease Study (CIDR) genotyped on the Illumina Omni1-Quad platform, and HRS control samples genotyped on the Illumina OMNI 2.5 platform. We also added data from HS SLE cases and controls genotyped on the Illumina HumanOmni 1 platform for a separate

GWAS study (unpublished data). Genotypes were merged with our data for further QC (described subsequently) to insure high-quality imputation.

Population structure and sample quality control

In order to obtain a homogeneous sample, for EA we removed individuals with <90% European ancestry. For AA, we removed individuals with <25% African ancestry. All participants in this study were filtered by PCA to identify population outliers as previously reported (88). Samples exceeding three standard deviations along any statistically significant principal component were defined as outliers and removed from the study. For each population, we merged genotype data with unrelated HapMap reference population data. We selected LD pruned variants ($r^2 < 0.2$) that had minor allele frequency (MAF) >1% and that were separated by at least 500 kilobases (kb). For PCA, we included 7170 SNPs for EA, 3501 SNPs for AA, 3646 SNPs for HS and 6928 SNPs for KR (Supplementary Material, Fig. S2). The number of SNPs chosen was a compromise between the size of the common data set and the number of independent SNPs. Although the number of SNPs to estimate population structure and ancestry proportions seems small, it is comparable to the sets used in other studies (89–90).

We used GCTA software (35) to estimate the first ten principal components for each ethnic population. All samples were plotted by their estimated ancestry proportion relative to the three continental reference populations (Supplementary Material, Fig. S3). We estimated the mean and standard deviation (SD) for each ethnic population for the first three principal components and removed samples that were outside three SD from the mean at each tail of the distribution.

To test for differences across data sets and potential population structure within each ethnic group, we selected variants from the combined data set common between unrelated HapMap samples (CEU, CHB + JPT and YRI) genotyped on the Illumina HumanOmni 2.5 platform. We selected LD pruned variants that had an MAF >1% and that were separated by at least 500 kb for estimation of admixture proportions. Ancestry proportion was estimated for all samples using ADMIXTURE (91) with 1–9 ancestry components. The optimum number of ancestry components was determined using the Bayesian Information Criterion (BIC) and the cross-validation error rate.

Imputation

Imputation is a statistical method used to probabilistically determine missing or untyped genotypes using a densely mapped reference panel. We performed imputation-based analysis in our ethnic-specific case-control samples. Genotypes included 145 SNPs from the ImmunoChip, 55 SNPs from EA and AA GWAS controls (34 overlapped between the platforms), and 19 SNPs from HS GWAS controls (all of which overlapped with ImmunoChip SNPs). We used publicly available 1000 Genomes Project data (2010-11-23 1000G Phase I) as reference panels for imputation, including 246 for AA (YRI + LWK + ASW), 381 for EA (GBR + FIN + IBS + CEU + TSI), 181 for HS (PUR + CLM + MXL) and 286 for Asians (AS; JPT + CHB + CHS). After QC 250 SNPs from the *NCF2* region (Chr1: 183,523,470–183,573,858 bp) were available for

analysis. Imputation was performed using MACH-ADMIX (31), which provided an imputation quality metric (Rsq). To insure high-quality imputation, only SNPs that passed post-imputation QC ($Rsq > 0.7$, $P_{HWE} > 0.001$ in controls, $MAF > 0.5\%$) were used for analysis.

Threshold for population-specific gene-wide significance adjusting for multiple testing and LD

Since we used a dense set of SNPs within *NCF2* to correct for multiple testing in each population, we used the spectral decomposition method implemented in SNPSpD (32,33). For each population, the experiment-wide significance level ($\alpha = 0.05$) was adjusted for multiple testing by estimating the effective number of independent SNPs by spectral decomposition of the pair-wise LD matrices adjusted with a Bonferroni correction. Since LD structures were different between the four populations, the effective numbers of independent SNPs (capturing most of the information across the entire *NCF2* gene) were also different across populations.

Association analysis, population stratification adjustment and conditional analysis

We performed single SNP association analysis in each population to compare allele frequencies between cases and controls using the χ^2 test (one degree of freedom), and calculated odds ratios (OR) and 95% confidence intervals (95% CI) using the PLINK (30) software package. In order to account for population structure, we used the first three principal components identified through PCA (92) as covariates in a logistic regression model. We also verified our results separately by adjusting three of the four individual ancestry proportions estimated by ADMIXTURE (93) as covariates. By using both types of stratification correction alternatively, we ensured that we accounted for stratification.

We used several methods to identify SNPs with independent effects in each ethnic population. First, we performed pair-wise conditional analysis in a step-wise manner using WHAP (94). Initially, we conditioned on the most significant SNP, then on the next most significant SNP, and so on, for each SNP that had passed the population-specific significance threshold in the imputation-based analysis. At each step, SNPs that lost significance after conditioning on the lead SNP were excluded from subsequent analyses until all residual association within the gene was accounted for. Secondly, we confirmed independent association using haplotype analysis implemented in PLINK (Supplementary Material, Table S5). Both omnibus and haplotype-specific association tests were used to evaluate these independent SNPs. An omnibus test was used to assess all haplotypes combined for SLE-association. Once association was observed with the omnibus test, a SNP-specific test as well as haplotype-specific test were used to identify whether any specific SNP or haplotype containing risk allele(s) explaining the omnibus association. Thirdly, we also verified independence between SNPs by their LD (r^2).

SNP–SNP interaction

For genetically complex diseases such as SLE, multiple interacting loci may contribute to disease susceptibility. Therefore, we

assessed SNP–SNP interaction (epistasis) to better understand *NCF2* association with SLE. First, we examined SNP–SNP interactions for independently associated SNPs within each population using a logistic regression model implemented in R. To account for population structure, we also adjusted for either the first three principal components estimated by GCTA or by major individual ancestry populations estimated by ADMIXTURE, described earlier. Significance of the relevant model terms were evaluated by comparing the log-likelihood difference between the null model (only independent effects) and the full model (with SNP–SNP interaction) with a χ^2 distribution with degrees of freedom (df) equal to the difference in the numbers of unconstrained parameters between models. We confirmed our results using the logistic regression model implemented in FITF (36), a PLINK (30) two-locus test and with GAIA (95).

Secondly, we verified our earlier results of potential SNP–SNP interaction using non-parametric generalized multifactor dimensionality reduction (GMDR) technique (96,97). A 10-fold cross-validation that randomly divides each data set into 10 randomly selected independent data sets (nine training sets and one testing set). This approach is the gold-standard in data mining and minimizes the bias introduced (due to random variation in the split or overestimation of the results) by averaging results across many divisions of the data (Dr Jason H. Moore, personal communication). The epistasis model (and its prediction error) was estimated for each training set and was then validated (98). The greatest cross-validation consistency (CVC) and the highest prediction accuracy were selected as the best predictor of this disease outcome. *P*-values were determined by the sign test, a robust non-parametric test implemented in the software.

Once we detected significant interaction between a pair of SNPs, we then used χ^2 test to assess the magnitude of association between individuals with at least one risk allele at either independently associated SNP, or at both independently associated SNPs, compared with those with no risk alleles as a baseline. ORs and *P*-values were used to assess the magnitude of the effect size.

In silico expression analysis

We obtained data from GeneVar (37,38) (Array Express Accession E-MTAB-264, protocol A-MEXP-930) for EBV-transformed lymphoblastoid cells from four HapMap populations (CEU = 35, CHB + JPT = 80, YRI = 38). Expression quantitative trait locus (eQTL; Illumina Human WG-6) data were merged with the corresponding genotypes at the independently associated variants identified in this study. Association analysis for *cis*-eQTLs in *NCF2* (GI_4557786-S) and neighboring genes *SMG7* (GI_42476073-A and GI_42475557-I) and *ARPC5* (GI_13569955-S and GI_23238212-S) were performed in PLINK. To confirm these results, we used data from the mRNA by SNP browser (39,40) to investigate whether any of these SNPs were associated with changes in transcript abundance. The mRNA by SNP browser data were obtained from lymphoblastoid cells of 206 families of British descent (297 sibling pairs and 11 half-sibling pairs) using the Affymetrix HG-U133 Plus 2.0 Chip (39,40). Association between SNPs and expression probes in the mRNA browser was determined

by regression analysis as described in Dixon *et al.* (40). We identified tagSNPs in linkage disequilibrium ($r^2 \geq 0.2$) with independently associated SNPs in CEU, YRI and CHB + JPT populations as a proxy for EA, AA and KR, respectively.

Luciferase reporter assay

We cloned exonic 106-mer sequence (bases 183,542,269–183,542,375) containing risk or non-risk alleles of rs17849501 into a luciferase assay vector, under a minimal thymidylate kinase promoter (TK_{min}, Xactagen Inc., USA). MonoMac-6 cells were transiently transfected in one of four groups: vector only, vector carrying multiple cloning sites (MCS), risk allele carrying DNA or non-risk allele carrying DNA sequences. Luciferase expression was measured at 6 h post-transfection as suggested (99), using the Gaussia Luciferase Assay (GLOW) with a Gaussia luciferase assay kit (Xactagen, LLC). Media was aspirated from transfected cells, which were plated and washed once. One hundred microliters of GLum.1 assay buffer (plus co-lenterazine substrate) and 50 μ l of GLum.1 assay reagent were added to each well and dark incubated for 5 min prior to reading luminescence (SpectraMax L, Molecular Devices: 475 nm). Measurements (relative luminescence units) were normalized to untransfected cells as controls. After normalization, statistical significance was calculated using *t*-tests assuming unequal variance.

Bioinformatics

All chromatin properties were examined using the UCSC Genome Browser (<http://genome.ucsc.edu>), based on the ENCODE annotations and other data contained in all 'Regulation' tracks.

Protein modeling

Structure figures were generated using PyMOL (<http://www.pymol.org>). The structure of the first SH3 domain (rs35937854/V297A) was taken from PDB file 2DMO. The structure of the PB1 domain (rs13306575/R395W and rs17849502/H389Q) was taken from PDB file 1OEY (44).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are grateful to the affected and unaffected individuals who participated in this study. We thank the research assistants, coordinators, and physicians who helped in the recruitment of subjects, including the individuals in the coordinating projects. We are indebted to J. Donald Capra, M.D for critical reading and helpful comments on this manuscript.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by grants from the US National Institutes of Health (AI103399, AR060366, AI094377, AI083194, CA141700, AR058621, AI101934, AI082714, AR053483, GM103510, GM103456) and A121983, the Korea Healthcare Technology R&D Project, Ministry for Health & Welfare, Republic of Korea.

REFERENCES

- Lawrence, R.C., Helmick, C.G., Arnett, F.C., Deyo, R.A., Felson, D.T., Giannini, E.H., Heyse, S.P., Hirsch, R., Hochberg, M.C., Hunder, G.G. *et al.* (1998) Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis Rheum.*, **41**, 778–799.
- Cervera, R., Khamashta, M.A., Font, J., Sebastiani, G.D., Gil, A., Lavilla, P., Mejia, J.C., Aydintug, A.O., Chwalinska-Sadowska, H., de Ramon, E. *et al.* (2003) Morbidity and mortality in systemic lupus erythematosus during a 10-year period: a comparison of early and late manifestations in a cohort of 1,000 patients. *Medicine (Baltimore)*, **82**, 299–308.
- Duffy, K.N., Duffy, C.M. and Gladman, D.D. (1991) Infection and disease activity in systemic lupus erythematosus: a review of hospitalized patients. *J. Rheumatol.*, **18**, 1180–1184.
- Souza, D.C., Santo, A.H. and Sato, E.I. (2012) Mortality profile related to systemic lupus erythematosus: a multiple cause-of-death analysis. *J. Rheumatol.*, **39**, 496–503.
- Alarcon, G.S., McGwin, G. Jr., Bastian, H.M., Roseman, J., Lisse, J., Fessler, B.J., Friedman, A.W. and Reveille, J.D. (2001) Systemic lupus erythematosus in three ethnic groups. VII [correction of VIII]. Predictors of early mortality in the LUMINA cohort. LUMINA Study Group. *Arthritis Rheum.*, **45**, 191–202.
- Goldblatt, F., Chambers, S., Rahman, A. and Isenberg, D.A. (2009) Serious infections in British patients with systemic lupus erythematosus: hospitalisations and mortality. *Lupus*, **18**, 682–689.
- Noel, V., Lortholary, O., Casassus, P., Cohen, P., Genereau, T., Andre, M.H., Mouthon, L. and Guillemin, L. (2001) Risk factors and prognostic influence of infection in a single cohort of 87 adults with systemic lupus erythematosus. *Ann. Rheum. Dis.*, **60**, 1141–1144.
- Cervera, R., Khamashta, M.A., Shoenfeld, Y., Camps, M.T., Jacobsen, S., Kiss, E., Zeher, M.M., Tincani, A., Kontopoulou-Griva, I., Galeazzi, M. *et al.* (2008) Morbidity and mortality in the antiphospholipid syndrome during a 5-year period: a multicenter prospective study of 1,000 patients. *Ann. Rheum. Dis.*, **68**, 1428–1432.
- Souza, D.C., Santo, A.H. and Sato, E.I. (2010) Trends in systemic lupus erythematosus mortality rates in the state of Sao Paulo, Brazil from 1985 to 2004. *Clin. Exp. Rheumatol.*, **28**, 519–524.
- Kundu, S., Ghosh, P., Datta, S., Ghosh, A., Chattopadhyay, S. and Chatterjee, M. (2012) Oxidative stress as a potential biomarker for determining disease activity in patients with rheumatoid arthritis. *Free Radic. Res.*, **46**, 1482–1489.
- Wang, G., Li, H. and Firoze Khan, M. (2012) Differential oxidative modification of proteins in MRL+/+ and MRL/lpr mice: increased formation of lipid peroxidation-derived aldehyde-protein adducts may contribute to accelerated onset of autoimmune response. *Free Radic. Res.*, **46**, 1472–1481.
- Winyard, P.G., Ryan, B., Eggleton, P., Nissim, A., Taylor, E., Lo Faro, M.L., Burkholz, T., Szabo-Taylor, K.E., Fox, B., Viner, N. *et al.* (2011) Measurement and meaning of markers of reactive species of oxygen, nitrogen and sulfur in healthy human subjects and patients with inflammatory joint disease. *Biochem. Soc. Trans.*, **39**, 1226–1232.
- Liu, Q., Cheng, L.I., Yi, L., Zhu, N., Wood, A., Changprirao, C.M., Ward, J.M. and Jackson, S.H. (2009) P47phox deficiency induces macrophage dysfunction resulting in progressive crystalline macrophage pneumonia. *Am. J. Pathol.*, **174**, 153–163.
- Yamauchi, A., Kim, C., Li, S., Marchal, C.C., Towe, J., Atkinson, S.J. and Dinayer, M.C. (2004) Rac2-deficient murine macrophages have selective defects in superoxide production and phagocytosis of opsonized particles. *J. Immunol.*, **173**, 5971–5979.
- Palmby, T.R., Abe, K., Karnoub, A.E. and Der, C.J. (2004) Vav transformation requires activation of multiple GTPases and regulation of gene expression. *Mol. Cancer Res.*, **2**, 702–711.

16. Gateva, V., Sandling, J.K., Hom, G., Taylor, K.E., Chung, S.A., Sun, X., Ortmann, W., Kosoy, R., Ferreira, R.C., Nordmark, G. *et al.* (2009) A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat. Genet.*, **Nov**, 1228–1233.
17. Cunninghame Graham, D.S., Morris, D.L., Bhangale, T.R., Criswell, L.A., Syvanen, A.C., Ronnblom, L., Behrens, T.W., Graham, R.R. and Vyse, T.J. (2011) Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS Genet.*, **7**, e1002341.
18. Jacob, C.O., Eisenstein, M., Dinauer, M.C., Ming, W., Liu, Q., John, S., Quismorio, F.P. Jr., Reiff, A., Myones, B.L., Kaufman, K.M. *et al.* (2012) Lupus-associated causal mutation in neutrophil cytosolic factor 2 (NCF2) brings unique insights to the structure and function of NADPH oxidase. *Proc. Natl Acad. Sci. USA*, **109**, E59–E67.
19. Yu, B., Chen, Y., Wu, Q., Li, P., Shao, Y., Zhang, J., Zhong, Q., Peng, X., Yang, H., Hu, X. *et al.* (2011) The association between single-nucleotide polymorphisms of NCF2 and systemic lupus erythematosus in Chinese mainland population. *Clin. Rheumatol.*, **30**, 521–527.
20. Deng, Y. and Tsao, B.P. (2010) Genetic susceptibility to systemic lupus erythematosus in the genomic era. *Nat. Rev. Rheumatol.*, **6**, 683–692.
21. Harley, I.T., Kaufman, K.M., Langefeld, C.D., Harley, J.B. and Kelly, J.A. (2009) Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. *Nat. Rev. Genet.*, **10**, 285–290.
22. Petri, M., Perez-Guthann, S., Longenecker, J.C. and Hochberg, M. (1991) Morbidity of systemic lupus erythematosus: role of race and socioeconomic status. *Am. J. Med.*, **91**, 345–353.
23. Reveille, J.D., Moulds, J.M., Ahn, C., Friedman, A.W., Baethge, B., Roseman, J., Straaton, K.V. and Alarcon, G.S. (1998) Systemic lupus erythematosus in three ethnic groups: I. The effects of HLA class II, C4, and CR1 alleles, socioeconomic factors, and ethnicity at disease onset. LUMINA Study Group. Lupus in minority populations, nature versus nurture. *Arthritis Rheum.*, **41**, 1161–1172.
24. Fernandez, M., Alarcon, G.S., Calvo-Alen, J., Andrade, R., McGwin, G. Jr., Vila, L.M., Reveille, J.D. and Group, L.S. (2007) A multiethnic, multicenter cohort of patients with systemic lupus erythematosus (SLE) as a model for the study of ethnic disparities in SLE. *Arthritis Rheum.*, **57**, 576–584.
25. Manson, J.J. and Rahman, A. (2006) Systemic lupus erythematosus. *Orphanet J. Rare Dis.*, **1**, 6.
26. Hochberg, M.C. (1987) The application of genetic epidemiology to systemic lupus erythematosus. *J. Rheumatol.*, **14**, 867–869.
27. Tan, E.M., Cohen, A.S., Fries, J.F., Masi, A.T., McShane, D.J., Rothfield, N.F., Schaller, J.G., Talal, N. and Winchester, R.J. (1982) The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.*, **25**, 1271–1277.
28. Cortes, A. and Brown, M.A. (2011) Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.*, **13**, 101.
29. Mailman, M.D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L. *et al.* (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–1186.
30. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
31. Liu, E.Y., Li, M., Wang, W. and Li, Y. (2013) MaCH-admix: genotype imputation for admixed populations. *Genet. Epidemiol.*, **37**, 25–37.
32. Nyholt, D.R. (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.*, **74**, 765–769.
33. Li, J. and Ji, L. (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)*, **95**, 221–227.
34. Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.*, **19**, 1655–1664.
35. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
36. Millstein, J., Conti, D.V., Gilliland, F.D. and Gauderman, W.J. (2006) A testing framework for identifying susceptibility genes in the presence of epistasis. *Am. J. Hum. Genet.*, **78**, 15–27.
37. Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P. and Dermitzakis, E.T. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*, **26**, 2474–2476.
38. Stranger, B.E., Montgomery, S.B., Dimas, A.S., Parts, L., Stegle, O., Ingle, C.E., Sekowska, M., Smith, G.D., Evans, D., Gutierrez-Arcelus, M. *et al.* (2012) Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.*, **8**, e1002639.
39. Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E. *et al.* (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*, **448**, 470–473.
40. Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
41. Unterholzner, L. and Izaurralde, E. (2004) SMG7 Acts as a molecular link between mRNA surveillance and mRNA decay. *Mol. Cell*, **16**, 587–596.
42. Bachmann, M.P., Bartsch, H., Gross, J.K., Maier, S.M., Gross, T.F., Workman, J.L., James, J.A., Farris, A.D., Jung, B., Franke, C. *et al.* (2006) Autoimmunity as a result of escape from RNA surveillance. *J. Immunol.*, **177**, 1698–1707.
43. Zhao, J., Sakurai, D., Deng, Y., Sestak, A.L., Langefeld, C.D., Kaufman, K.M. and Kelly, J.A. (2011) In *American College of Rheumatology. Arthritis Rheum.*, Chicago, Illinois, Vol. 63 Suppl 10, pp. 639.
44. Wilson, M.I., Gill, D.J., Perisic, O., Quinn, M.T. and Williams, R.L. (2003) PB1 domain-mediated heterodimerization in NADPH oxidase and signaling complexes of atypical protein kinase C with Par6 and p62. *Mol. Cell*, **12**, 39–50.
45. Yanagisawa, H., Nagasawa, T., Kuramochi, S., Abe, T., Ikawa, Y. and Todokoro, K. (1991) Constitutive expression of exogenous c-myc gene causes maturation block in monocyte-macrophage differentiation. *Biochim. Biophys. Acta*, **1088**, 380–384.
46. Maehara, Y., Miyano, K. and Sumimoto, H. (2009) Role for the first SH3 domain of p67phox in activation of superoxide-producing NADPH oxidases. *Biochem. Biophys. Res. Commun.*, **379**, 589–593.
47. Ming, W., Li, S., Billadeau, D.D., Quilliam, L.A. and Dinauer, M.C. (2007) The Rac effector p67phox regulates phagocyte NADPH oxidase by stimulating Vav1 guanine nucleotide exchange activity. *Mol. Cell Biol*, **27**, 312–323.
48. Song, E., Jaishankar, G.B., Saleh, H., Jithpratuck, W., Sahni, R. and Krishnaswamy, G. (2011) Chronic granulomatous disease: a review of the infectious and inflammatory complications. *Clin. Mol. Allergy*, **9**, 10.
49. Roesler, J., Segerer, F., Morbach, H., Kleinert, S., Thieme, S., Rosen-Wolff, A. and Liese, J.G. (2012) P67-phox (NCF2) lacking exons 11 and 12 is functionally active and leads to an extremely late diagnosis of chronic granulomatous disease (CGD). *PLoS One*, **7**, e34296.
50. Grizot, S., Fieschi, F., Dagher, M.C. and Pebay-Peyroula, E. (2001) The active N-terminal region of p67phox. Structure at 1.8 Å resolution and biochemical characterizations of the A128 V mutant implicated in chronic granulomatous disease. *J. Biol. Chem.*, **276**, 21627–21631.
51. Mizuki, K., Takeya, R., Kuribayashi, F., Nobuhisa, I., Kohda, D., Nunoi, H., Takeshige, K. and Sumimoto, H. (2005) A region C-terminal to the proline-rich core of p47phox regulates activation of the phagocyte NADPH oxidase by interacting with the C-terminal SH3 domain of p67phox. *Arch. Biochem. Biophys.*, **444**, 185–194.
52. de Boer, M., Hartl, D., Wintergerst, U., Belohradsky, B.H. and Roos, D. (2005) A donor splice site mutation in intron 1 of CYBA, leading to chronic granulomatous disease. *Blood Cells Mol. Dis.*, **35**, 365–369.
53. Jirapongsananuruk, O., Noack, D., Boonchoo, S., Thepthai, C., Chokephaibulkit, K., Visitsunthorn, N., Vichyanond, P., Luangwedchakarn, V., Likasitwattanukul, S. and Piboonpocanun, S. (2007) A novel mutation of the CYBB gene resulting in severe form of X-linked chronic granulomatous disease. *Asian Pac. J. Allergy Immunol.*, **25**, 249–252.
54. Lewis, E.M., Singla, M., Sergeant, S., Koty, P.P. and McPhail, L.C. (2008) X-linked chronic granulomatous disease secondary to skewed X chromosome inactivation in a female with a novel CYBB mutation and late presentation. *Clin. Immunol.*, **129**, 372–380.
55. Teimourian, S., Zomorodian, E., Badalzadeh, M., Pouya, A., Kannengiesser, C., Mansouri, D., Cheraghi, T. and Parvaneh, N. (2008) Characterization of six novel mutations in CYBA: the gene causing autosomal recessive chronic granulomatous disease. *British J. Haematol.*, **141**, 848–851.
56. El Kares, R., Barbouche, M.R., Elloumi-Zghal, H., Bejaoui, M., Chemli, J., Mellouli, F., Tebib, N., Abdelmoula, M.S., Boukthir, S., Fitouri, Z. *et al.* (2006) Genetic and mutational heterogeneity of autosomal recessive chronic granulomatous disease in Tunisia. *J. Hum. Genet.*, **51**, 887–895.
57. van den Berg, J.M., van Koppen, E., Ahlin, A., Belohradsky, B.H., Bernatowska, E., Corbeel, L., Espanol, T., Fischer, A., Kurenko-Deptuch,

- M., Mouy, R. *et al.* (2009) Chronic granulomatous disease: the European experience. *PLoS One*, **4**, e5234.
58. Koker, M.Y., van Leeuwen, K., de Boer, M., Celmeli, F., Metin, A., Ozgur, T.T., Tezcan, I., Sanal, O. and Roos, D. (2009) Six different CYBA mutations including three novel mutations in ten families from Turkey, resulting in autosomal recessive chronic granulomatous disease. *Eur. J. Clin. Invest.*, **39**, 311–319.
 59. Gentsch, M., Kaczmarczyk, A., van Leeuwen, K., de Boer, M., Kaus-Drobek, M., Dagher, M.C., Kaiser, P., Arkwright, P.D., Gahr, M., Rosen-Wolff, A. *et al.* (2010) Alu-repeat-induced deletions within the NCF2 gene causing p67-phox-deficient chronic granulomatous disease (CGD). *Hum. Mutat.*, **31**, 151–158.
 60. Roos, D., Kuhns, D.B., Maddalena, A., Roesler, J., Lopez, J.A., Ariga, T., Avcin, T., de Boer, M., Bustamante, J., Condino-Neto, A. *et al.* (2010) Hematologically important mutations: X-linked chronic granulomatous disease (third update). *Blood Cells Mol. Dis.*, **45**, 246–265.
 61. Teimourian, S., de Boer, M. and Roos, D. (2010) Molecular basis of autosomal recessive chronic granulomatous disease in Iran. *J. Clin. Immunol.*, **30**, 587–592.
 62. Roos, D., Kuhns, D.B., Maddalena, A., Bustamante, J., Kannengiesser, C., de Boer, M., van Leeuwen, K., Koker, M.Y., Wolach, B., Roesler, J. *et al.* (2010) Hematologically important mutations: the autosomal recessive forms of chronic granulomatous disease (second update). *Blood Cells Mol. Dis.*, **44**, 291–299.
 63. Noack, D., Rae, J., Cross, A.R., Munoz, J., Salmen, S., Mendoza, J.A., Rossi, N., Curnutte, J.T. and Heyworth, P.G. (1999) Autosomal recessive chronic granulomatous disease caused by novel mutations in NCF-2, the gene encoding the p67-phox component of phagocyte NADPH oxidase. *Hum. Genet.*, **105**, 460–467.
 64. Patino, P.J., Rae, J., Noack, D., Erickson, R., Ding, J., de Olarte, D.G. and Curnutte, J.T. (1999) Molecular characterization of autosomal recessive chronic granulomatous disease caused by a defect of the nicotinamide adenine dinucleotide phosphate (reduced form) oxidase component p67-phox. *Blood*, **94**, 2505–2514.
 65. Lapouge, K., Smith, S.J., Walker, P.A., Gamblin, S.J., Smerdon, S.J. and Rittinger, K. (2000) Structure of the TPR domain of p67phox in complex with Rac.GTP. *Mol. Cell*, **6**, 899–907.
 66. Muike, A.M., Xu, W., Guo, C.H., Walters, T.D., Wolters, V.M., Fattouh, R., Lam, G.Y., Hu, P., Murchie, R., Sherlock, M. *et al.* (2011) NADPH oxidase complex and IBD candidate gene studies: identification of a rare variant in NCF2 that results in reduced binding to RAC2. *Gut*, **61**, 1028–1035.
 67. Roberts, R.L., Hollis-Moffatt, J.E., Gearry, R.B., Kennedy, M.A., Barclay, M.L. and Merriman, T.R. (2008) Confirmation of association of IRGM and NCF4 with ileal Crohn's disease in a population-based cohort. *Genes Immun.*, **9**, 561–565.
 68. Somasundaram, R., Deuring, J.J., van der Woude, C.J., Peppelenbosch, M.P. and Fuhler, G.M. (2012) Linking risk conferring mutations in NCF4 to functional consequences in Crohn's disease. *Gut*, **61**, 1097; author reply 1097–1098.
 69. Olsson, L.M., Lindqvist, A.K., Kallberg, H., Padyukov, L., Burkhardt, H., Alfredsson, L., Klareskog, L. and Holmdahl, R. (2007) A case-control study of rheumatoid arthritis identifies an associated single nucleotide polymorphism in the NCF4 gene, supporting a role for the NADPH-oxidase complex in autoimmunity. *Arthritis Res. Ther.*, **9**, R98.
 70. Hunt, K.A., Mistry, V., Bockett, N.A., Ahmad, T., Ban, M., Barker, J.N., Barrett, J.C., Blackburn, H., Brand, O., Burren, O. *et al.* (2013) Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature*, **498**, 232–235.
 71. Babior, B.M. (2004) NADPH oxidase. *Curr. Opin. Immunol.*, **16**, 42–47.
 72. Bylund, J., Brown, K.L., Movitz, C., Dahlgren, C. and Karlsson, A. (2010) Intracellular generation of superoxide by the phagocyte NADPH oxidase: how, where, and what for? *Free Radic. Biol. Med.*, **49**, 1834–1845.
 73. Nauseef, W.M. (2008) Biological roles for the NOX family NADPH oxidases. *J. Biol. Chem.*, **283**, 16961–16965.
 74. Bae, Y.S., Oh, H., Rhee, S.G. and Yoo, Y.D. (2011) Regulation of reactive oxygen species generation in cell signaling. *Mol. Cells*, **32**, 491–509.
 75. de Oliveira-Junior, E.B., Bustamante, J., Newburger, P.E. and Condino-Neto, A. (2011) The human NADPH oxidase: primary and secondary defects impairing the respiratory burst function and the microbicidal ability of phagocytes. *Scand. J. Immunol.*, **73**, 420–427.
 76. Ferro, E., Goitre, L., Retta, S.F. and Trabalzini, L. (2012) The interplay between ROS and Ras GTPases: physiological and pathological implications. *J. Signal. Transduct.*, **2012**, 365769.
 77. Bhogal, R.H., Weston, C.J., Curbishley, S.M., Adams, D.H. and Afford, S.C. (2012) Autophagy: a cyto-protective mechanism which prevents primary human hepatocyte apoptosis during oxidative stress. *Autophagy*, **8**, 545–558.
 78. Lande, R., Ganguly, D., Facchinetti, V., Frasca, L., Conrad, C., Gregorio, J., Meller, S., Chamilos, G., Sebasigari, R., Riccieri, V. *et al.* (2011) Neutrophils activate plasmacytoid dendritic cells by releasing self-DNA-peptide complexes in systemic lupus erythematosus. *Sci. Transl. Med.*, **3**, 73ra19.
 79. Kirchner, T., Moller, S., Klinger, M., Solbach, W., Laskay, T. and Behnen, M. (2012) The impact of various reactive oxygen species on the formation of neutrophil extracellular traps. *Mediators Inflamm.*, **2012**, 849136.
 80. Campbell, A.M., Kashgarian, M. and Shlomchik, M.J. (2012) NADPH oxidase inhibits the pathogenesis of systemic lupus erythematosus. *Sci. Transl. Med.*, **4**, 157ra141.
 81. Elkon, K.B. and Santer, D.M. (2012) Complement, interferon and lupus. *Curr. Opin. Immunol.*, **24**, 665–670.
 82. Bouts, Y.M., Wolthuis, D.F., Dirkx, M.F., Pieterse, E., Simons, E.M., van Boekel, A.M., Dieker, J.W. and van der Vlag, J. (2012) Apoptosis and NET formation in the pathogenesis of SLE. *Autoimmunity*, **45**, 597–601.
 83. Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D.S., Weinrauch, Y. and Zychlinsky, A. (2004) Neutrophil extracellular traps kill bacteria. *Science*, **303**, 1532–1535.
 84. Arazna, M., Pruchniak, M.P., Zycinska, K. and Demkow, U. (2013) Neutrophil extracellular trap in human diseases. *Adv. Exp. Med. Biol.*, **756**, 1–8.
 85. Leffler, J., Martin, M., Gullstrand, B., Tyden, H., Lood, C., Truedsson, L., Bengtsson, A.A. and Blom, A.M. (2012) Neutrophil extracellular traps that are not degraded in systemic lupus erythematosus activate complement exacerbating the disease. *J. Immunol.*, **188**, 3522–3531.
 86. Rasmussen, A., Sevier, S., Kelly, J.A., Glenn, S.B., Aberle, T., Cooney, C.M., Grether, A., James, E., Ning, J., Tesiram, J. *et al.* (2011) The Lupus Family Registry and Repository. *Rheumatology (Oxford)*, **50**, 47–59.
 87. Hochberg, M.C. (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.*, **40**, 1725.
 88. Adrianto, I., Wen, F., Templeton, A., Wiley, G., King, J.B., Lessard, C.J., Bates, J.S., Hu, Y., Kelly, J.A., Kaufman, K.M. *et al.* (2011) Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nat. Genet.*, **43**, 253–258.
 89. Liu, J.Z., Hov, J.R., Folseraas, T., Ellinghaus, E., Rushbrook, S.M., Doncheva, N.T., Andreassen, O.A., Weersma, R.K., Weismuller, T.J., Eksteen, B. *et al.* (2013) Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat. Genet.*, **45**, 670–675.
 90. Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhenakova, A., Stahl, E., Viatte, S., McAllister, K. *et al.* (2012) High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.*, **44**, 1336–1340.
 91. Patterson, N., Hattangadi, N., Lane, B., Lohmueller, K.E., Hafler, D.A., Oksenberg, J.R., Hauser, S.L., Smith, M.W., O'Brien, S.J., Altshuler, D. *et al.* (2004) Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.*, **74**, 979–1000.
 92. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
 93. Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.*, **19**, 1655–1664.
 94. Purcell, S., Daly, M.J. and Sham, P.C. (2007) WHAP: haplotype-based association analysis. *Bioinformatics*, **23**, 255–256.
 95. Macgregor, S. and Khan, I.A. (2006) GAIA: an easy-to-use web-based application for interaction analysis of case-control data. *BMC Med. Genet.*, **7**, 34.
 96. Chen, G.B., Xu, Y., Xu, H.M., Li, M.D., Zhu, J. and Lou, X.Y. (2011) Practical and theoretical considerations in study design for detecting gene-gene interactions using MDR and GMDR approaches. *PLoS One*, **6**, e16981.
 97. Lou, X.Y., Chen, G.B., Yan, L., Ma, J.Z., Zhu, J., Elston, R.C. and Li, M.D. (2007) A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am. J. Hum. Genet.*, **80**, 1125–1137.
 98. Ritchie, M.D., Hahn, L.W., Roodi, N., Bailey, L.R., Dupont, W.D., Parl, F.F. and Moore, J.H. (2001) Multifactor-dimensionality reduction reveals

high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.*, **69**, 138–147.

99. Klan, N. and Steinhilber, D. (2003) Transient transfection of the human myeloid cell line Mono Mac 6 using electroporation. *Biotechniques*, **34**, 142–147.

APPENDIX

We are also grateful for the assistance provided by the GENLES Network group: Eduardo Acevedo, MD (Hospital Nacional Guillermo Almenara Irigoyen, Lima, Perú), Ignacio García-De La Torre, MD (Hospital General de Occidente, Secretaría de Salud, Zapopan, Jalisco, México), Marco A. Maradiaga-Ceceña, MD (Hospital General de Culiacán, Culiacán, Mexico), Mario H. Cardiel, MD, MSc (Hospital General Dr Miguel Silva, Morelia, Mexico), Jorge A. Esquivel-Valerio, MD (Servicio de Reumatología, Hospital Universitario Dr José Eleuterio González de la Universidad Autónoma de Nuevo Leon, Monterrey, Nuevo León, Mexico), Jacqueline Rodriguez-Amado, MD (Servicio de Reumatología,

Hospital Universitario Dr José Eleuterio González de la Universidad Autónoma de Nuevo Leon, Monterrey, Nuevo León, Mexico), José Francisco Moctezuma, MD (Hospital General de México, Mexico City, Mexico), Pedro Miranda, MD (Servicio de Reumatología, Hospital San Juan de Dios, Santiago, Chile), Carlos Perandones, MD (Centro de Educación Médica e Investigaciones Clínicas (CEMIC), Buenos Aires, Argentina), Cecilia Castel, MD (Servicio de Inmunología, Hospital Central de Mendoza, Mendoza, Argentina), Hugo A. Laborde, MD (Servicio de Reumatología, Hospital de Clínicas 'José de San Martín', Buenos Aires, Argentina), Paula Alba, MD (Laboratorio de Inmunología, Hospital Córdoba, Córdoba, Argentina), Jorge Musuruana, MD (Servicio de Reumatología, Hospital José Bernardo Iturraspe, Santa Fé, Argentina), Annelise Goecke, MD (Hospital Clínico, Universidad de Chile, Santiago de Chile, Chile), Carola Foster, MD (Laboratorio de Reumatología, Hospital Del Salvador, Santiago de Chile, Chile), Lorena Orozco (Instituto Nacional de Medicina Genómica, Mexico City, Mexico), Vicente Baca (Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico).