Title

A missense variant in NCF1 is associated with susceptibility to multiple autoimmune diseases

Permalink

https://escholarship.org/uc/item/5134j5gx

Journal

Nature Genetics, 49(3)

ISSN

1061-4036

Authors

Zhao, Jian
Ma, Jianyang
Deng, Yun
et al.

Publication Date

2017-03-01

DOI

10.1038/ng.3782

Peer reviewed

Published in final edited form as:
Nat Genet. 2017 March ; 49(3): 433–437. doi:10.1038/ng.3782.

A missense variant in *NCF1* is associated with susceptibility to multiple autoimmune diseases

Jian Zhao[1,2,15,*], **Jianyang Ma**[3,15], **Yun Deng**[1,2], **Jennifer A. Kelly**[4], **Kwangwoo Kim**[5], **So-Young Bang**[6], **Hye-Soon Lee**[6], **Quan-Zhen Li**[7], **Edward K. Wakeland**[7], **Rong Qiu**[8], **Mengru Liu**[9], **Jianping Guo**[9], **Zhanguo Li**[9], **Wenfeng Tan**[10], **Astrid Rasmussen**[4], **Christopher J. Lessard**[4,11], **Kathy L. Sivils**[4,11], **Bevra H. Hahn**[2], **Jennifer M. Grossman**[2], **Diane L. Kamen**[1], **Gary S. Gilkeson**[1], **Sang-Cheol Bae**[6], **Patrick M. Gaffney**[4], **Nan Shen**[3,8,12,13,14,*], and **Betty P. Tsao**[1,2,*]

[1]Division of Rheumatology and Immunology, Department of Medicine, Medical University of South Carolina, Charleston, South Carolina 29425, USA

[2]Division of Rheumatology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA

[3]Shanghai Institute of Rheumatology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200001, China

[4]Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104, USA

[5]Department of Biology, Kyung Hee University, Seoul 02447, Korea

[6]Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul 04763, Korea

[7]Department of Immunology, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

---

[*] Correspondence should be addressed to B.P.T. (tsaob@musc.edu), J.Z. (zhaoji@musc.edu) or N.S. (nanshensibs@gmail.com); Betty P. Tsao, PhD, 936C Clinical Science Bldg, MUSC, 96 Jonathan Lucas St, Charleston, SC 29425, USA, tsaob@musc.edu, Phone: 843-792-7966, Fax: 843-792-2562; Jian Zhao, PhD, 916C Clinical Science Bldg, MUSC, 96 Jonathan Lucas St, Charleston, SC 29425, USA, zhaoji@musc.edu, Phone: 843-792-5667, Fax: 843-792-2562; Nan Shen, MD, 320 Yueyang Road Bldg A1307, Institute of Health Sciences, Shanghai 200031, China, nanshensibs@gmail.com, Phone: 86-21-63260477, Fax: 86-21-63363475.

[15] These authors contributed equally to this work.

Author Contributions

J.Z., B.P.T. and N.S. led the study. J.Z., Y.D. and B.P.T. wrote the paper. J.Z., J.M., Y.D., and R.Q. performed the experiments. J.Z., J.M., Y.D., J.A.K and K.K. analyzed the data and performed statistical analysis. S.Y.B, H.S.L, Q.Z.L., E.K.W., M.L., J.G., Z.L., W.T., A.R., C.J.L., K.L.S., B.H.H., J.M.G., D.L.K., G.S.G., S.C.B. and P.M.G. contributed primarily to sample collection and/or genotyping. All authors reviewed the final manuscript.

Competing Financial Interests Statement

The authors declare no competing financial interests.

URLs:

1000 Genomes Project, http://www.1000genomes.org
RegulomeDB, http://www.regulomedb.org
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink
SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html
HAPLOVIEW, https://www.broadinstitute.org/haploview/haploview
EIGENSOFT, https://www.hsph.harvard.edu/alkes-price/software
UCSC genome browser, http://genome.ucsc.edu

[8]Institute of Health Sciences, Shanghai Jiao Tong University School of Medicine and Shanghai Institutes for Biological Sciences (SIBS), University of Chinese Academy of Sciences, Chinese Academy of Sciences (CAS), Shanghai, China

[9]Department of Rheumatology and Immunology, Peking University People's Hospital, Beijing 100044, China

[10]Department of Rheumatology, the First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210005, China

[11]Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, USA

[12]State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200001, China

[13]The Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA

[14]Collaborative Innovation Center for Translational Medicine at Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

Abstract

Systemic lupus erythematosus (SLE) is a heterogeneous autoimmune disease with a strong genetic component characterized by autoantibody production and a type I interferon signature[1]. Here we report a missense variant (g.74779296G>A; p.Arg90His) in *NCF1*, encoding the p47^{phox} subunit of the phagocyte NADPH oxidase (NOX2), as the putative underlying causal variant that drives a strong SLE-associated signal detected by the Immunochip in the *GTF2IRD1-GTF2I* region at 7q11.23 with a complex genomic structure. We show that the p.Arg90His substitution, which is reported to cause reduced reactive oxygen species (ROS) production[2], predisposes to SLE (odds ratio (OR)=3.47 in Asians (P_{meta}=3.1×10^{-104}), OR=2.61 in European Americans, OR=2.02 in African Americans) and other autoimmune diseases, including primary Sjögren's syndrome (OR=2.45 in Chinese, OR=2.35 in European Americans) and rheumatoid arthritis (OR=1.65 in Koreans). Additionally, decreased and increased copy numbers of *NCF1* predispose to and protect against SLE, respectively. Our data highlight the pathogenic role of reduced NOX2-derived ROS levels in autoimmune diseases.

Dozens of SLE-associated loci have been identified by genome-wide association studies (GWAS) and included on the ImmunoChip for fine-mapping[3,4]. Using the ImmunoChip, we genotyped DNA samples from SLE cases and healthy controls from Chinese, European-American and African-American ancestry groups. In Chinese, the strongest association signal was detected at rs73366469 (minor allele frequency of 28.7% in cases versus 12.6% in controls, P=3.8×10^{-29}, OR=2.88) within the *GTF2IRD1-GTF2I* intergenic region at 7q11.23 rather than SLE-associated GWAS loci (Fig.1a), consistent with another Asian Immunochip study[5]. This association was confirmed in European Americans at a modest significance level (P=7.5×10^{-3}, OR=1.32), but not in African Americans (Supplementary Table 1).

Because rs73366469 has no strong functional implication (RegulomeDB score = 5) and is not in linkage disequilibrium (LD; $r^2<0.04$) with any Immunochip SNP, we hypothesized that the underlying causal variant(s) were not on the Immunochip and were in strong, modest and weak LD with rs73366469 in Chinese, European Americans and African Americans, respectively. In the 1000 Genomes Project data set, we found two non-Immunochip SNPs (rs117026326 and rs12667901) that were in stronger LD with rs73366469 in Asians than in Europeans and Africans (Fig.1 and Supplementary Table 2). In a subset of Chinese and European-American subjects genotyped using the ImmunoChip (Chinese-1 and EurAm-1) and independent replication data sets (Chinese-2 and EurAm-2), rs117026326, located in intron 9 of *GTF2I*, exhibited stronger association with SLE ($P_{meta\ Chinese}$=4.4×10^{-40}, OR=2.94; $P_{meta\ EurAm}$=1.2×10^{-4}, OR=2.83, Supplementary Table 3 and 4) than rs73366469 and rs12667901, and conditioning on rs117026326 rather than rs73366469 or rs12667901 eliminated association signals for the other two SNPs, suggesting that association of rs73366469 with SLE might be attributed to rs117026326. However, rs117026326 was nearly non-polymorphic in African Americans and was not associated with SLE (Supplementary Table 5), which explained the lack of association between rs73366469 and SLE in African Americans.

Located 62kb from rs117026326 is *NCF1* (neutrophil cytosolic factor 1), encoding the regulatory p47^{phox} subunit necessary for activation of the phagocytic NOX2 complexl, which is likely an SLE-related gene because nonfunctional NOX2 exacerbates lupus features in lupus-prone MRL.*Fas*^{lpr} mice and induces lupus-like type I interferon signature, autoantibody production and immune complex deposition in kidneys of BALB/c.*Ncf1*^{m1J} mice[6,7]; *NCF2*, encoding another NOX2 regulatory subunit, p67^{phox}, harbors the missense variant p.His389Gln that is associated with SLE risk in European Americans[8]. However, rs117026326 genotype was not associated with the transcript levels of *NCF1* or other two neighboring genes, *GTF2I* and *GTF2IRD1*, in the peripheral blood mononuclear cells (PBMCs) of patients with SLE and healthy controls (Supplementary Fig. 1). Notably, the *NCF1* region is barely covered by variants in the 1000 Genomes Project phase 1 data set (Fig.1a), probably because *NCF1* has 98% sequence identity with the nonfunctional pseudogenes *NCF1B* and *NCF1C* at 7q11.23 that are located within large DNA duplications[9] (Fig.1b and 2a). The strong association of rs117026326 with SLE and functional implication of *NCF1* led us to further hypothesize that rs117026326 might tag causal variant(s) of *NCF1* not present in the 1000 Genomes Project.

The GTGT sequence within exon 2 of *NCF1* is a well-characterized variant that distinguishes this gene from *NCF1B and NCF1C*, which contain a GT deletion (ΔGT)[10] (Fig.2b). However, reciprocal crossover results in the presence of GTGT-containing *NCF1B and NCF1C* and ΔGT-containing *NCF1*[11,12] (Fig.3a). To obtain the correct genotypes for *NCF1* variants, we specifically amplified the *NCF1* sequence using PCR targeting GTGT Supplementary Fig.2) and measured the ΔGT/GTGT ratio by real-time PCR to exclude subjects carrying GTGT-containing *NCF1B and NCF1C* (3:3 or 2:4 ratios) or ΔGT-containing *NCF1* (5:1 ratio) (Fig.3b and c). Using this approach followed by Sanger sequencing, we resequenced the entire 15.5-kb *NCF1* region in 45 Chinese subjects and identified 67 SNPs (Supplementary Table 6). Of these SNPs, only two encoding p.Arg90His (rs201802880) and p.Ser99Gly (rs17295741) in exon 4 and two intronic SNPs (intronic-1 (without a dbSNP id) and intronic-2 (rs199789198)) of *NCF1* showed r^2>0.1 with rs117026326. We hypothesized that at least one of these four *NCF1* SNPs was the causal variant tagged by rs117026326 in Chinese and shared by European and African Americans and that the causal SNP(s) might be located by leveraging the different LD patterns in these ancestry groups.

LD analysis in African Americans suggested that either p.Arg90His or p.Ser99Gly might be the causal variant because intronic-1 and intronic-2 were in complete LD with non-polymorphic rs117026326 (Fig.1c). To confirm this, we assessed these four variants, by performing nested PCR and Taqman assays and measuring the ΔGT/GTGT ratio, for association with SLE in African Americans. Of the variants, only p.Arg90His was associated with SLE (15.7% versus 8.3%, P=2.9×10^{-5}, OR=2.02, Table 1 and Supplementary Table 5). As expected, intronic-1 and intronic-2 were nearly non-polymorphic and were not associated with SLE in African Americans, suggesting that they might not be the causal variant shared by different ancestry groups, and they were excluded in subsequent analyses.

Next, we assessed p.Arg90His and p.Ser99Gly for association with SLE in Asians and European Americans. In two Chinese and one Korean data set, p.Arg90His exhibited stronger association with SLE (Chinese-1: 38.1% versus 15.6%, P=2.6×10^{-23}, OR=3.35; Chinese-2: 41.6% versus 16.7%, P=1.5×10^{-41}, OR=3.27; Korean: 46.6% versus 18.1%, P=2.6×10^{-43}, OR=3.82; $P_{\text{meta-Asian}}$=3.1×10^{-104}, OR=3.47, Table 1 and Supplementary Table 3) than p.Ser99Gly, and conditioning on p.Arg90His eliminated association signals detected at p.Ser99Glyand the *GTF2IRD1-GTF2I* region. In two European-Americandata sets, p.Arg90His consistently showed stronger association with SLE (EurAm-1: 5.5% versus 2.1%, P=9.5×10^{-5}, OR=2.83; EurAm-2: 6.0% versus 2.4%, P=5.0×10^{-4}, OR=2.42; $P_{\text{meta-EurAm}}$=1.9×10^{-7}, OR=2.61, Table 1 and Supplementary Table 4) than p.Ser99Glyand explained association signals at p.Ser99Gly and the *GTF2IRD1-GTF2I* region in a conditional test. Furthermore, the allele encoding p.Arg90Hiswas dose dependently associated with early age of onset in Korean and European-American patients with SLE (Supplementary Fig.3). These data support p.Arg90His as a likely causal variant for SLE susceptibility shared across Asian, European-American and African-American populations within the *GTF2IRD1-GTF2I-NCF1* region.

In addition to SLE, p.Arg90His was associated with other autoimmune diseases, including primary Sjögren's syndrome ([MIM 270150]) in Chinese (37.8% versus 18.3%, P=7.2×10^{-17}, OR=2.45) and European Americans (4.8% versus 2.2%, P=9.7×10^{-4}, OR=2.35) (Table 1 and Supplementary Table 7), which explained the reported association of rs117026326 with Sjögren's syndrome in Chinese[13], and seropositive rheumatoid arthritis ([MIM 180300]) in Koreans (26.6% versus 18.1%, P=2.5×10^{-8}, OR=1.65, Table 1 and Supplementary Table 8) but with a modest effect size.

Arg90 of $p47^{phox}$, located in a phosphoinositide-binding pocket of the PX domain, has a crucial role in the membrane translocation of cytosolic $p47^{phox}$ and resultant activation of NOX2 for ROS production[14]. The substitution of evolutionarily conserved Arg90 with a histidine residue encoded by the SLE risk allele was predicted to be deleterious (Supplementary Fig.4), a prediction that is supported by *in vitro* mutation studies indicating that changing Arg90 to histindine[2], or lysine, leucine or alanine[14–16], reduces ROS production.l These data suggested that p.Arg90His might confer risk for SLE by reducing NOX2-derived ROS levels. Consistent with this idea, the SLE risk allele encoding p.His389Gln in p67^{phox} causes reduced ROS production in transfection assay[8]. However, p.Arg90Hiswas not associated with intracellular ROS levels in neutrophils from controls (Supplementary Fig.5), probably because of the impact of mitochondrial ROS[17].

The GTGT to ΔGT mutation in *NCF1* (rs273585651) leads to a frameshift and a premature stop codon at residue 51 (Fig. 2b)[11]. Because of an absence of functional p47^{phox} and failure of ROS production, homozygous carriers of ΔGT in *NCF1* develop a rare disease, chronic granulomatous disease (CGD), and have increased risk of developing SLE (Fig. 3a)[7]. *NCF1B* and *NCF1C* are transcribed but do not produce functional protein because they contain ΔGT[10]. However, GTGT-containing *NCF1B* and *NCF1C* are believed to produce intact p47^{phox} similar to functional *NCF1*, although this has not yet been experimentally validated[12]. Analyzing the ΔGT/GTGT ratio allowed us to assess copy number variation (CNV) in *NCF1*. One copy of *NCF1* (5:1 ratio) was associated with increased risk of SLE in Koreans (P=0.032), Chinese (P=0.011) and European Americans ($P=5.9\times10^{-4}$, OR=3.91, Supplementary Table 9). In contrast, having ≥3 copies of *NCF1* (3:3 and 2:4 ratios) was protective against developing SLE in Koreans ($P=3.7\times10^{-5}$), Chinese ($P=2.8\times10^{-3}$, OR=0.28), European Americans (P=0.038, OR=0.85) and African Americans (P=0.018, OR=0.73). This data support the notion that reduced ROS production is a risk factor for SLE. Association between CNV of *NCF1* and Sjögren's syndrome was not detected in European Americans, probably owing to the limited sample size.

ROS can be a double-edged sword in autoimmunity. High levels of ROS, predominantly produced by NOX2 in phagocytes for host defense, may lead to inflammatory tissue damage, but ROS are also signaling molecules regulating T cell differentiation, B cell proliferation and antigen processing in dendritic cells[18]. Our findings suggest that reduced NOX2-derived ROS production increases the risk of developing autoimmune diseases, consistent with previous reports[2,6–8], but the underlying mechanism remains elusive. Of interest, NOX2-derived ROS are required for LC3-associated phagocytosis[19] and nonfunctional NOX2 causes defective clearance of dying cells and lupus-like phenotypes[20], providing a possible explanation for the pathogenic role of reduced ROS level in SLE.

Of note, *NCF1* variants have not been correctly called in studies using short sequence reads, such as the 1000 Genomes Project (Supplementary Fig.6) and the Exome Aggregation Consortium, owing to the presence of *NCF1B and NCF1C*. Consequently, we would like to emphasize that *NCF1* variants need to be assessed with great caution to exclude the impact of *NCF1B and NCF1C*. In this study, we identified p.Arg90His by *NCF1*-specific PCR and Sanger sequencing. However, because of the complexity of this region and the difficulty of long-range PCR, we only resequenced *NCF1* in Chinese subjects, and potential causal variants within the *NCF1*-neighboring region or specific for European and African Americans might not have been discovered. Given that most lupus-associated variants show OR<2 in GWAS[21], the chance of finding a causal variant within the noncoding region of *NCF1* that can explain the association of p.Arg90His with SLE (OR>3 in Asians and >2 in European and African Americans) is small. Although we cannot exclude the possibility of additional causal *NCF1* variants, their association with SLE should be independent from that of P.ARG90HIS.

In summary, we identify a p.Arg90His substitution encoded in *NCF1* as a novel risk variant for SLE, Sjögren's syndrome and rheumatoid arthritis, and we show that decreased and increased copy numbers of *NCF1* predispose to and protect against SLE, respectively. Our data highlight the pathogenic role of reduced ROS production in autoimmune diseases and indicate the presence of missing heritability within complex genomic regions.

Data Availability

Sequencing data and summary-level association data are shown in Supplementary Tables 1 and 3–9. Raw TaqMan data for p.Arg90His are shown as Supplementary Data.

Online Methods

Subjects

Discovery stage—Patients with SLE and healthy controls in the discovery stage were recruited from the University of California Los Angeles (UCLA), the Oklahoma medical research foundation (OMRF) and the Medical University of South Carolina (MUSC). All patients with SLE met at least 4 of the 11 American College of Rheumatology (ACR) criteria for the classification of SLE[22]. The final data set after quality control comprised subjects from three different ancestry groups, including Chinese (1010 cases and 848 controls recruited from UCLA), African Americans (532 cases and 367 controls recruited from OMRF and MUSC) and European Americans (930 cases and 1107 controls recruited from UCLA and OMRF). All these samples were genotyped using the Immunochip, and we estimated that there was power of >90% to identify an SLE-associated variant with MAF>10% and OR>2.0 at the GWAS significance level of $P<5\times10^{-8}$ in Chinese and European Americans. Of the subjects, all African Americans and a subset of the Chinese (441 cases and 589 controls; dataset Chinese-1] and European Americans (716 cases and 578 controls; dataset EurAm-1) subjects were genotyped using Taqman assays for SNPs not included on the Immunochip.

Replication stage—To replicate the result in data set Chinese-1, an independent cohort comprising 746 Chinese SLE cases and 1034 Chinese healthy controls (data set Chinese-2) was recruited from Shanghai Renji Hospital. In addition, another Asian replication cohort comprising 614 Korean SLE cases and 692 healthy Korean controls was recruited from the Hanyang University Hospital for Rheumatic Diseases (HUHRD). To replicate the result in data set EurAm-1, an independent European-American cohort comprising 875 SLE cases and 540 healthy controls (data set EurAm-2) were recruited from UCLA, MUSC and OMRF.

To assess p.Arg90His and p.Ser99Gly for association with other autoimmune diseases, 863 Korean patients with rheumatoid arthritis were recruited from HUHRD. All patients with rheumatoid arthritis were positive for antibodies to anticitrullinated peptide and fulfilled the ACR 1987 revised criteria for the classification of RA[23]. In addition, 382 European-

American patients with primary Sjögren's syndrome, described in a previous GWAS[24], were recruited from OMRF, and 449 Chinese patients with Sjögren's syndrome and 469 healthy Chinese controls were recruited from Peking University People's Hospital. All patients with Sjögren's syndrome fulfilled the American-European Consensus Group (AECG) criteria for primary Sjögren's syndrome[25].

To measure the ΔGT/GTGT ratio, additional Chinese (198 patients with SLE and 471 controls recruited from the First Affiliated Hospital of Nanjing Medical University), African Ameericans (184 patients with SLE and 39 controls) and Koreans (215 patients with SLE and 90 controls) were used, but these subjects were not genotyped for SNPs.

Each participating institution had institutional review board (IRB) approval to recruit subjects. All subjects provided written informed consent.

Immunochip genotyping and quality control

Samples in the discovery stage were genotyped using the Immunochip according to Illumina's protocols at the University of Texas Southwestern Medical Center, HudsonAlpha or OMRF, and all samples were reclustered for genotype calling as a single project at OMRF.

We excluded SNPs with a call rate <95% in cases or controls and removed samples with a SNP call rate <90%. SNPs were also excluded if they showed deviation from Hardy-Weinberg equilibrium (HWE) ($P_{\mathrm{HWE}}<0.001$ in controls, $P_{\mathrm{HWE}}<0.00001$ in cases) or they had significant different call rates in cases and controls (call rate<98% and $P<0.05$). On the basis of the remaining SNPs, we identified related samples (shared identity by descent [PI_HAT] > 0.25, estimated using PLINK v1.07) and samples showing mismatch between the reported and estimated sex and excluded them from subsequent analyses. To identify ancestry outliers, the remaining samples were assessed by principal-component analysis (implemented in EIGENSOFT 4.2) based on 7,500 randomly selected autosomal SNPs with MAF>1%, low LD ($r^2<0.1$ with each other) and no evidence of association with SLE ($P>0.01$), and the 1000 Genomes Project samples were used as reference populations (including 286 Asians [97 CHB, 100 CHS, 89 JPT], 379 Europeans [85 CEU, 98 TSI, 89 GBR, 93 FIN, 14 IBS] and 246 Africans [88 YRI, 61 ASW, 97 LWK]) (Supplementary Fig. 7). Outliers of each ancestry (>6 s.d. from the mean principal component) were excluded. Principal components showing significant difference between cases and controls were included as covariates in the association test.

The Korean samples were analyzed in an independent Immunochip study and a GWAS study[5,26]. Principal components for the Korean samples were obtained from these two studies.

NCF1-specific PCR for DNA sequencing

To identify *NCF1* variant(s) tagged by rs117026326, a total of 45 Chinese subjects, including 23 homozygous for the risk allele and 22 homozygos for the non-risk allele at rs117026326, were selected for sequencing. Among these subjects, there was probability of 90% and 64% of discovering *NCF1* variants with MAF of 5% and 1%, respectively. To distinguish *NCF1* from *NCF1B* and *NCF1C*, we confirmed that all 45 subjects showed a 4:2 ratio of ΔGT/GTGT and amplified *NCF1*-specific DNA sequence by three PCR reactions (Supplementary Fig.2) using the LongRange PCR Kit (206402, QIAGEN). The PCR reactions contained 200nM of each primer, 600μM dNTP, 2.75mM Mg^{2+}, 1X buffer and 1U Taq polymerase. PCR reactions were run on the SimpliAmp thermal cycler (Thermo Fisher Scientific) with the following condition: 3 min at 93°C followed by 50 cycles of 15 s at 93°C, 1 min at 62°C and 7 min at 68 ˚C. PCR products were sequenced on the 3730xl DNA Analyzer (Thermo Fisher Scientific). Primers are shown in Supplementary Table 10.

NCF1-specific PCR and Taqman genotyping

rs117026326, rs73366469 and rs12667901 were directly genotyped using TaqMan assays (Thermo Fisher Scientific). Because of the presence of *NCF1B* and *NCF1C*, genotypes of p.Arg90His, p.Ser99Gly, intronic-1 and intronic-2 in *NCF1* were obtained by nested PCR and TaqMan assays. We first PCR amplified an *NCF1*-specific fragment by targeting the GTGT sequence in exon 2 of *NCF1* (P2 or P2*, as shown in Supplementary Fig.2), and each PCR fragment was subjected to agarose gel electrophoresis to assess the quality of specific amplification. Next, 1μl of the PCR fragment diluted 1:10,000 was used as the template in a TaqMan assay for SNP genotyping. Samples that failed in the first PCR reaction were removed from TaqMan analysis. Taqman assays were run on either ABI 7900HT or the QuantStudio™ 6 Flex RT-PCR System (Thermo Fisher Scientific). DNA sequences for TaqMan assays are shown in Supplementary Table 11. Raw TaqMan data of P.ARG90HIS are shown as Supplementary Data.

Determination of the ΔGT/GTGT ratio

To exclude subjects carrying GTGT-containing *NCF1B and NCF1C* or ΔGT-containing *NCF1* and determine the CNV of *NCF1*, *NCF1B and NCF1C*, we measured the ΔGT/GTGT ratio using Taqman CNV assays described in previous studies[2,12] (Fig.3 and Supplementary Table 11). In a duplex real time PCR reaction, the Taqman assay targeting either ΔGT or GTGT was run simultaneously with a copy number reference assay (RNase P) targeting the riboluclease P RNA component H1 gene (*RPPH1*) known to exist in two copies in the genome (4403326, Thermo Fisher Scientific). Each reaction was run in quadruplicate on the QuantStudio™ 6 Flex RT-PCR System (Thermo Fisher Scientific). The copy numbers of ΔGT and GTGT were compared with that of RNase P and calculated by the comparative C_t method using CopyCaller Software v2.0 (Thermo Fisher Scientific). We combined all subjects (n=6,914) for multiplate analysis and selected the "without calibrator sample" option in CopyCaller by assuming four copies of ΔGT and two copies of GTGT as the most frequent copy numbers. Ratios between the copy number of ΔGT and GTGT were normally distributed around 5, 2, 1 and 0.5 (Fig.3b). According to the distribution, subjects showing a ΔGT/GTGT ratio of >4, 1.3–2.6, 0.7–1.2 and <0.7 were assigned a theoretical ratio of 5:1, 4:2, 3:3 and 2:4, respectively (Fig.3b and c).

Association analyses

All African-American subjects (Supplementary Table 5), all European-American subjects (Supplementary Table 4 and 7), all Korean patients with SLE and controls (Supplementary Table 3) and a subset of subjects in the Chinese-1 data set (Supplementary Table 3) were assessed for the ΔGT/GTGT ratio. Association analyses were carried out either in subjects with the normal ΔGT/GTGT ratio of 4:2 (carrying 2 copies each of *NCF1*, *NCF1B* and *NCF1C*) or in all subjects. Given that the 5:1, 3:3 and 2:4 ratios were found in less than 2% of Asians (Fig.3c and Supplementary Table 9) and that their impact on association analyses of p.Arg90His in Asian data sets was negligible, Korean patients with rheumatoid arthritis (Supplementary Table 8) and some of the Chinese subjects (Supplementary Table 3 and 7) were not assessed for the ΔGT/GTGT ratio.

In each ancestry group, SNPs were assessed for association with disease using an additive model in logistic regression in which principal components showing significant differences between cases and controls (available for Chinese, European Americans and African Americans in the discovery stage and Koreans in the replication stage) and sex were included as covariates. Haplotype-based conditional association tests were performed to detect independent association signals, and meta-analysis was conducted to combine multiple datasets. All analyses described above were performed using PLINK v1.07. In addition, we calculated the Bayes factor and posterior probability for each SNP using SNPTEST v2.5.2. Pairwise LD values were calculated using Haploview 4.2. CNV of *NCF1* was assessed for association with SLE and Sjögren's syndrome using Fisher's exact test, and the 4:2 ratio of ΔGT/GTGT was used as reference genotype (OR=1).

Quantitative real-time PCR

Total mRNAs were extracted from peripheral blood mononuclear cells of SLE patients and controls using the All Prep DNA/RNA mini kit (QIAGEN) and then were reverse-transcribed into cDNAs (Thermo Fisher Scientific). Transcript levels of *NCF1*, *GTF2I*, *GTF2IRD1* and *GAPDH* were measured by quantitative real-time PCR using TaqMan assays (Hs00165362_m1, Hs01073660_m1, Hs00249456_m1 and Hs03929097_g1, Thermo Fisher Scientific, respectively). The relative expression levels of *NCF1*, *GTF2I* and *GTF2IRD1*, normalized to those of the housekeeping gene *GAPDH*, were calculated by the comparative C_t methodl.

Measurement of ROS levels in neutrophils

We measured *ex vivo* ROS levels in neutrophils for association with p.Arg90His genotypes. Fresh blood samples collected in vacutainer tubes containing EDTA were obtained from healthy Chinese subjects (n=101) recruited at Shanghai Renji Hospital. Neutrophils were isolated from blood by density gradient centrifugation using PolymorphPrep (Axis-Shield) and cultured at 37°C in RPMI-1640 medium (11875-093, Gibco) supplemented with 10% FBS. ROS levels were determined using DCFH-DA dye (S0033, Beyotime), which can be oxidized to fluorescent DCF by intracellular peroxides. Two million neutrophils were incubated with 5μM DCFH-DA for 20min and then stimulated with 30 ng/ml phorbol myristate acetate (PMA) (P1585, Sigma) for 1h. Cells were washed twice with PBS and assayed for the mean fluorescence intensity using flow cytometry (BD Biosciences). Data were processed using the FlowJo software. DNA samples were extracted using the TIANamp Blood DNA kit (DP348-03, TIANGEN) and assessed for p.Arg90Hisgenotype.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all subjects for participation in this study. We thank E. Magdangal and Y. Shi for help with DNA preparation and organization. We also thank A. Lusis for valuable discussion and comments. This work was supported by the US National Institutes of Health R01AR043814 (B.P.T.), R21AR065626 (B.P.T.), R01AR056360 (P.M.G.), R01AR063124 (P.M.G.), R01AR043274 (K.L.S.), R01DE015223 (K.L.S.), R01DE018209 (K.L.S.), R01AR050782 (K.L.S.), R01AR065953 (C.J.L. and K.L.S.), P50AR0608040 (K.L.S. and C.J.L.), U19AI082714 (P.M.G., K.L.S. and C.J.L.), P60AR062755 (D.L.K. and G.S.G.), P30AR053484 (Judith A. James), P30GM103510 (Judith A. James) and U54GM104938 (Judith A. James), the Lupus Foundation of America (B.P.T.), the Alliance for Lupus Research (B.P.T.), the Sjögren's Syndrome Foundation (K.L.S. and C.J.L.), the Korea Healthcare Technology R&D Project of Ministry for Health and Welfare in Republic of Korea HI13C2124 (S.C.B.) and HI15C3182 (K.K.), National Basic research Program of China (973 program) grant 2014CB541902 (N.S.), Key Research Program of Bureau of Frontier Sciences and Education Chinese Academy of Sciences grant QYZDJ-SSW-SMC006 (N.S.), Key Research Program of the Chinese Academy of Sciences grant KJZD-EW-L01-3 (N.S.), State Key Laboratory of Oncogenes and Related Genes grant 91-14-05 (N.S.), National Natural Science Foundation of China grant 31630021 (N.S.), Strategic Priority Research Program of the Chinese Academy of Sciences grant XDA12020107 (N.S.) and the Clinical and Translational Science Institute (CTSI) UL1RR033176 (UCLA), UL1TR000124 (UCLA) and UL1TR001450 (MUSC) and funds from the Spaulding-Paolozzi Autoimmunity Center of Excellence (MUSC), Richard M. Silver, MD, Endowment for Inflammation Research (B.P.T.) and SmartState® Center of Economic Excellence in Inflammation and Fibrosis Research (B.P.T.). The funders had no role in study design, data collection, analysis, and interpretation, writing of the report, or decision to submit the paper for publication.

References

1. Tsokos GC. Systemic lupus erythematosus. N Engl J Med. 2011; 365:2110–21. [PubMed: 22129255]
2. Olsson LM, et al. Copy number variation of the gene NCF1 is associated with rheumatoid arthritis. Antioxid Redox Signal. 2012; 16:71–8. [PubMed: 21728841]
3. Deng Y, Tsao BP. Advances in lupus genetics and epigenetics. Curr Opin Rheumatol. 2014; 26:482–92. [PubMed: 25010439]
4. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011; 13:101. [PubMed: 21345260]
5. Sun C, et al. High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. Nat Genet. 2016; 48:323–30. [PubMed: 26808113]
6. Campbell AM, Kashgarian M, Shlomchik MJ. NADPH oxidase inhibits the pathogenesis of systemic lupus erythematosus. Sci Transl Med. 2012; 4:157ra141.
7. Kelkka T, et al. Reactive oxygen species deficiency induces autoimmunity with type 1 interferon signature. Antioxid Redox Signal. 2014; 21:2231–45. [PubMed: 24787605]
8. Jacob CO, et al. 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 U S A. 2012; 109:E59–67. [PubMed: 22203994]
9. Peoples R, et al. A physical map, including a BAC/PAC clone contig, of the Williams-Beuren syndrome–deletion region at 7q11.23. Am J Hum Genet. 2000; 66:47–68. [PubMed: 10631136]
10. Gorlach A, et al. A p47-phox pseudogene carries the most common mutation causing p47-phox-deficient chronic granulomatous disease. J Clin Invest. 1997; 100:1907–18. [PubMed: 9329953]
11. Casimir CM, et al. Autosomal recessive chronic granulomatous disease caused by deletion at a dinucleotide repeat. Proc Natl Acad Sci U S A. 1991; 88:2753–7. [PubMed: 2011585]
12. Heyworth PG, Noack D, Cross AR. Identification of a novel NCF-1 (p47-phox) pseudogene not containing the signature GT deletion: significance for A47 degrees chronic granulomatous disease carrier detection. Blood. 2002; 100:1845–51. [PubMed: 12176908]
13. Li Y, et al. A genome-wide association study in Han Chinese identifies a susceptibility locus for primary Sjogren's syndrome at 7q11.23. Nat Genet. 2013; 45:1361–5. [PubMed: 24097066]

14. Ago T, et al. Phosphorylation of p47phox directs phox homology domain from SH3 domain toward phosphoinositides, leading to phagocyte NADPH oxidase activation. Proc Natl Acad Sci U S A. 2003; 100:4474–9. [PubMed: 12672956]
15. Kanai F, et al. The PX domains of p47phox and p40phox bind to lipid products of PI(3)K. Nat Cell Biol. 2001; 3:675–8. [PubMed: 11433300]
16. Li XJ, Marchal CC, Stull ND, Stahelin RV, Dinauer MC. p47phox Phox homology domain regulates plasma membrane but not phagosome neutrophil NADPH oxidase activation. J Biol Chem. 2010; 285:35169–79. [PubMed: 20817944]
17. Lood C, et al. Neutrophil extracellular traps enriched in oxidized mitochondrial DNA are interferogenic and contribute to lupus-like disease. Nat Med. 2016; 22:146–53. [PubMed: 26779811]
18. Cachat J, Deffert C, Hugues S, Krause KH. Phagocyte NADPH oxidase and specific immunity. Clin Sci (Lond). 2015; 128:635–48. [PubMed: 25760962]
19. Martinez J, et al. Molecular characterization of LC3-associated phagocytosis reveals distinct roles for Rubicon, NOX2 and autophagy proteins. Nat Cell Biol. 2015; 17:893–906. [PubMed: 26098576]
20. Martinez J, et al. Noncanonical autophagy inhibits the autoinflammatory, lupus-like response to dying cells. Nature. 2016; 533:115–9. [PubMed: 27096368]
21. Morris DL, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. Nat Genet. 2016; 48:940–6. [PubMed: 27399966]
22. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997; 40:1725.
23. Arnett FC, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988; 31:315–24. [PubMed: 3358796]
24. Lessard CJ, et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren’s syndrome. Nat Genet. 2013; 45:1284–92. [PubMed: 24097067]
25. Vitali C, et al. Classification criteria for Sjogren’s syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis. 2002; 61:554–8. [PubMed: 12006334]
26. Kim K, et al. High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. Ann Rheum Dis. 2015; 74:e13. [PubMed: 24532676]

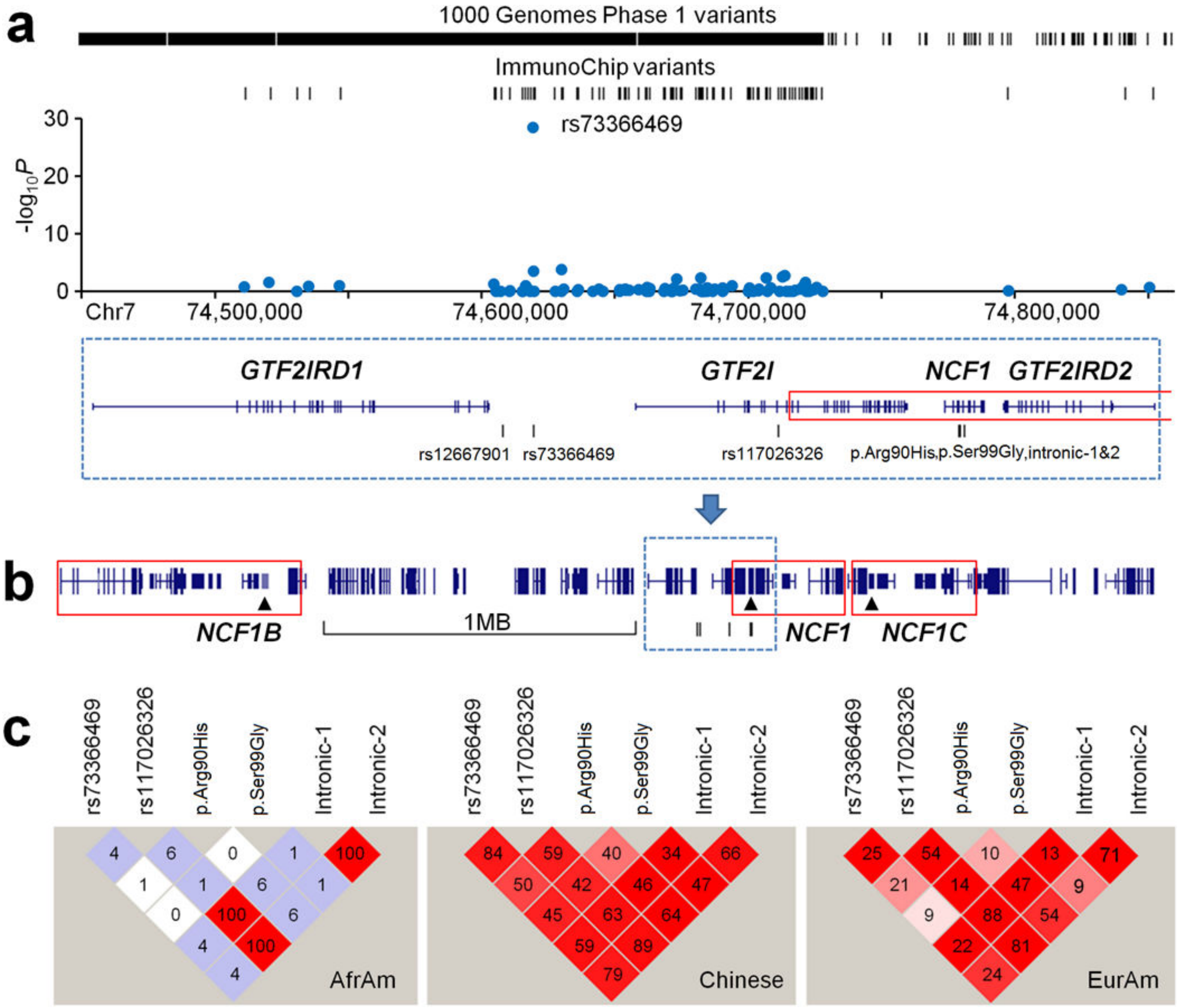

Fig 1. The *GTF2IRD1-GTF2I-NCF1* region on 7q11.23

(a) Association plot of Immunochip variants. The locations of 1000 Genomes Project variants are indicated on the top. The allelic *P* value ($-\log_{10}P$) of each Immunochip variant assessed for association with SLE in Chinese is plotted as a circle according to the location of the variant. A map of the SLE-associated SNPs described in this study is shown in a box outlined by a blue dash line. (b) Large duplications at 7q11.23 containing *NCF1*, *NCF1B* and *NCF1C*. Duplications are highlighted as red boxes in which the location of *NCF1*, *NCF1B* or *NCF1C* is indicated by a triangle. The region shown in panel a is highlighted by a blue dashed box. (c) LD (shown as r^2) analyses in African-American (AfrAm), Chinese and Eurorpean-American (EurAm) subjects with a 4:2 ratio of ΔGT/GTGT (n=100 for each ancestral group).

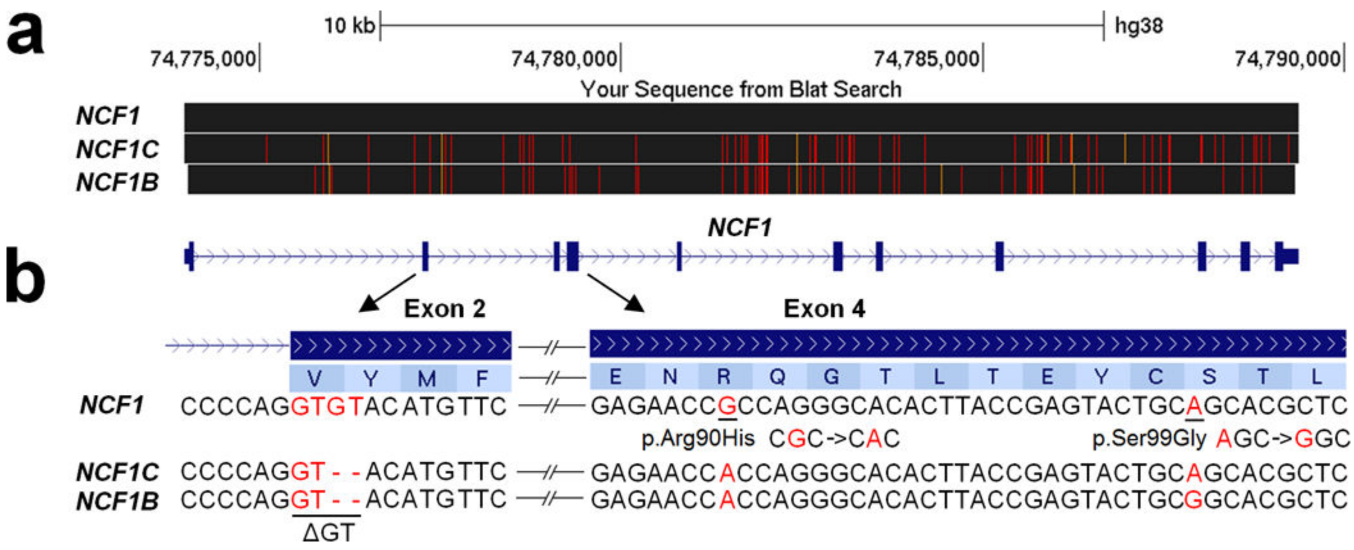

Fig 2. Highly homologous sequence among *NCF1*, *NCF1B* and *NCF1C*

(a) Alignment of the DNA sequences of *NCF1B* and *NCF1C* with that of *NCF1*. Sequences of *NCF1B* and *NCF1C* that are not identical to that of *NCF1* are shown in red. This figure was generated using the BLAT tool in the UCSC Genome Browser. (b) *NCF1* variants GTGT, p.Arg90His and p.Ser99Gly and the corresponding sequences at *NCF1B* and *NCF1C* (highlighted in red). Sequences are based on human reference genome build GRCh38/hg38.

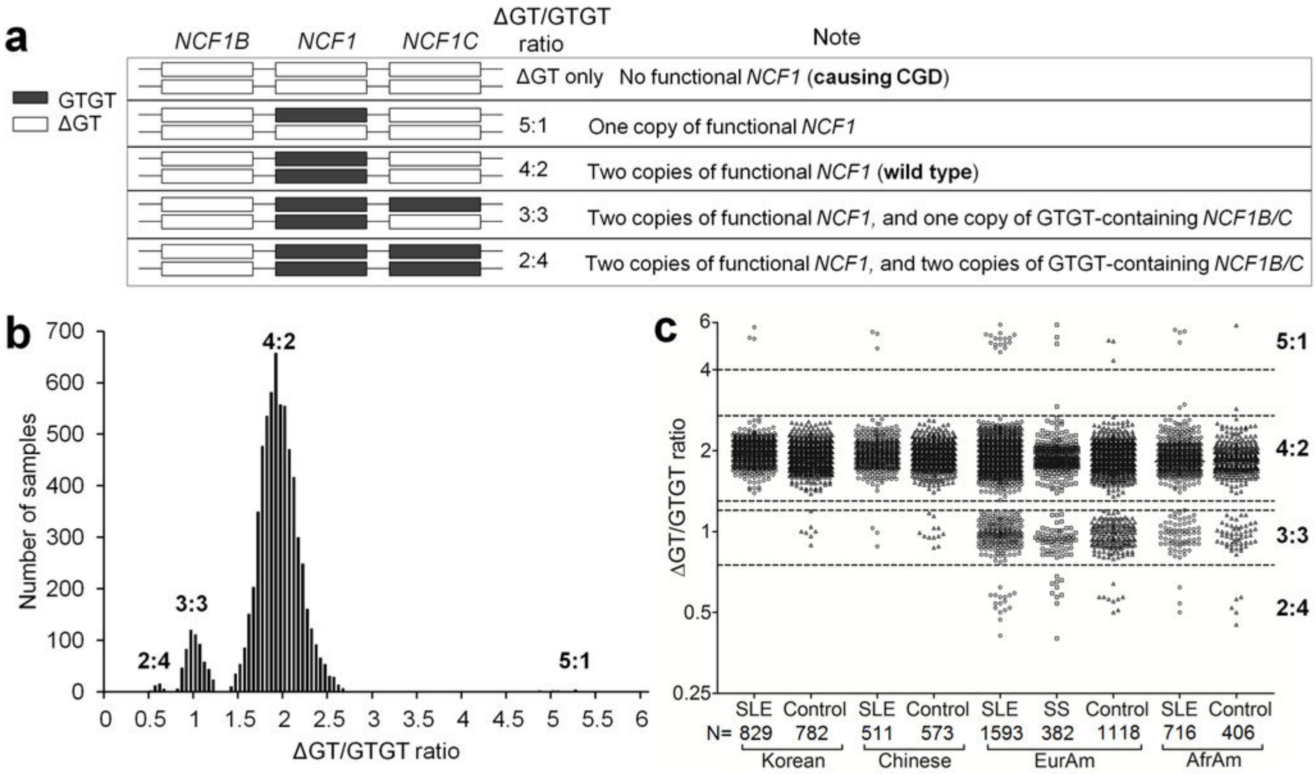

Fig 3. Determination of the ΔGT/GTGT ratio

(a) ΔGT/GTGT ratios. (b) Distribution of the ΔGT/GTGT ratio in all studied subjects. (c) Plot of the ΔGT/GTGT ratio in different ancestry groups.

Table 1

Association of SNPs in the *GTF2IRD1-GTF2I-NCF1* region with autoimmune diseases in different ancestral groups

				rs73366469 (C/T^a)					rs117026326 (T/C^a)					p.Arg90His (A/G^a)				p.Ser99Gly (G/A^a)				
		Subject		MAF^b (%)					MAF^b (%)					MAF^b (%)				MAF^b (%)				
Disease	**Dataset**	**Case**	**Control**	**Case**	**Control**	***P***	**OR**	***P^c***	**Case**	**Control**	***P***	**OR**	***P^c***	**Case**	**Control**	***P***	**OR**	**Case**	**Control**	***P***	**OR**	***P^c***
SLE	AfrAm*	474	315	18.5	19.5	0.66	0.94		0.43	0.0	1.00			15.7	8.3	2.9×10^{-5}	2.02	40.1	39.4	0.78	1.03	
SLE	Chinese-1	436	586	29.7	13.9	1.7×10^{-14}	2.66	0.57	28.1	11.4	1.8×10^{-17}	3.18	0.27	38.1	15.6	2.6×10^{-23}	3.35	43.2	26.8	1.6×10^{-13}	2.22	0.70
SLE	Chinese-2	746	1034	30.0	15.2	6.5×10^{-22}	2.48	0.046	27.1	11.9	2.3×10^{-24}	2.82	0.093	41.6	16.7	1.5×10^{-41}	3.27	46.5	28.2	1.6×10^{-22}	2.13	0.21
SLE	Meta-Chinese	1182	1620			9.2×10^{-35}	2.55				4.4×10^{-40}	2.94				4.2×10^{-63}	3.30			2.1×10^{-34}	2.16	
SLE	Korean*	611	685						30.0	11.4	3.3×10^{-27}	3.25	0.18	46.6	18.1	2.6×10^{-43}	3.82	52.4	30.5	3.2×10^{-26}	2.47	0.052
SLE	Meta-Asians	1793	2305								2.0×10^{-65}	3.05				3.1×10^{-104}	3.47			1.5×10^{-58}	2.26	
SLE	EurAm-1*	612	485	10.8	8.5	0.036	1.37	0.64	2.4	0.72	6.4×10^{-3}	3.25	0.59	5.5	2.1	9.5×10^{-5}	2.83	17.8	13.0	3.9×10^{-3}	1.44	0.12
SLE	EurAm-2*	736	446	11.9	8.5	0.012	1.46	0.25	2.4	0.67	6.1×10^{-3}	3.51	0.29	6.0	2.4	5.0×10^{-4}	2.42	18.6	13.1	5.5×10^{-4}	1.53	0.057
SLE	Meta-EurAm	1348	931			1.1×10^{-3}	1.42				1.1×10^{-4}	3.37				1.9×10^{-7}	2.61			7.3×10^{-6}	1.49	
SS	Chinese-SS	449	469						26.2	14.7	4.6×10^{-9}	2.03	0.29	37.8	18.3	7.2×10^{-17}	2.45	41.5	29.1	9.7×10^{-8}	1.68	0.017
SS	EurAm-SS*	302	931						2.0	0.70	7.3×10^{-3}	3.12	0.36	4.8	2.2	9.7×10^{-4}	2.35	15.4	13.1	0.20	1.19	0.93
RA	Korean-RA	836	685											26.6	18.1	2.5×10^{-8}	1.65	35.2	30.5	2.0×10^{-3}	1.27	0.94

[a] minor/major allele,

[b] minor allele frequency,

[c] *P* value after conditioning on p.Arg90His.

SLE, systemic lupus erythematosus; SS, Sjogren's Syndrome; RA, rheumatoid arthritis

[*] Association analyses were conducted in subjects with the 4:2 ratio of ΔGT/GTGT.