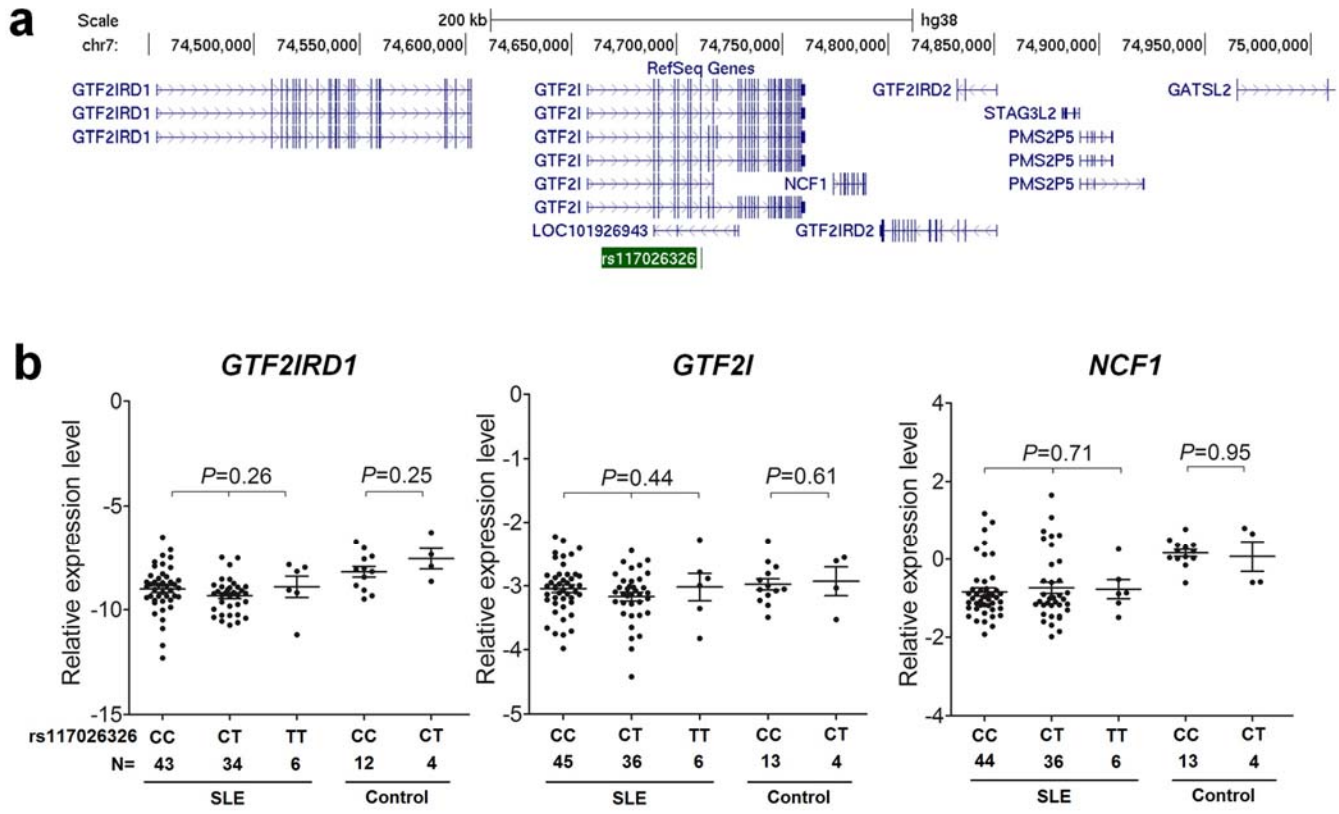
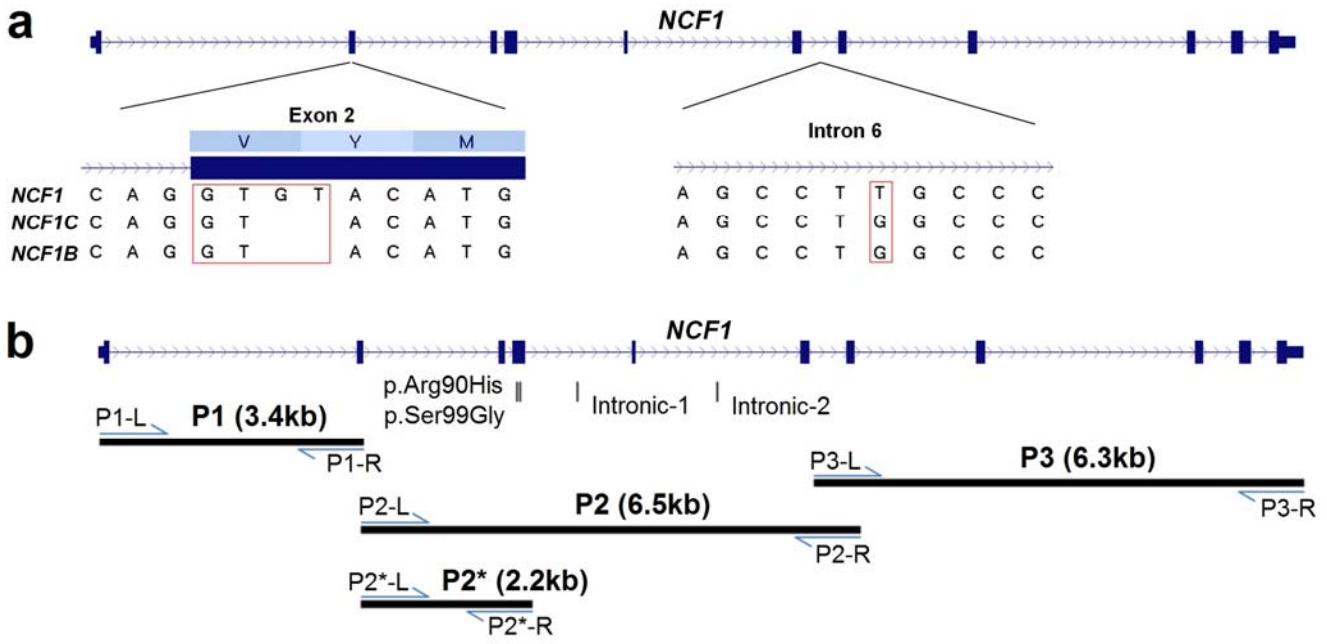
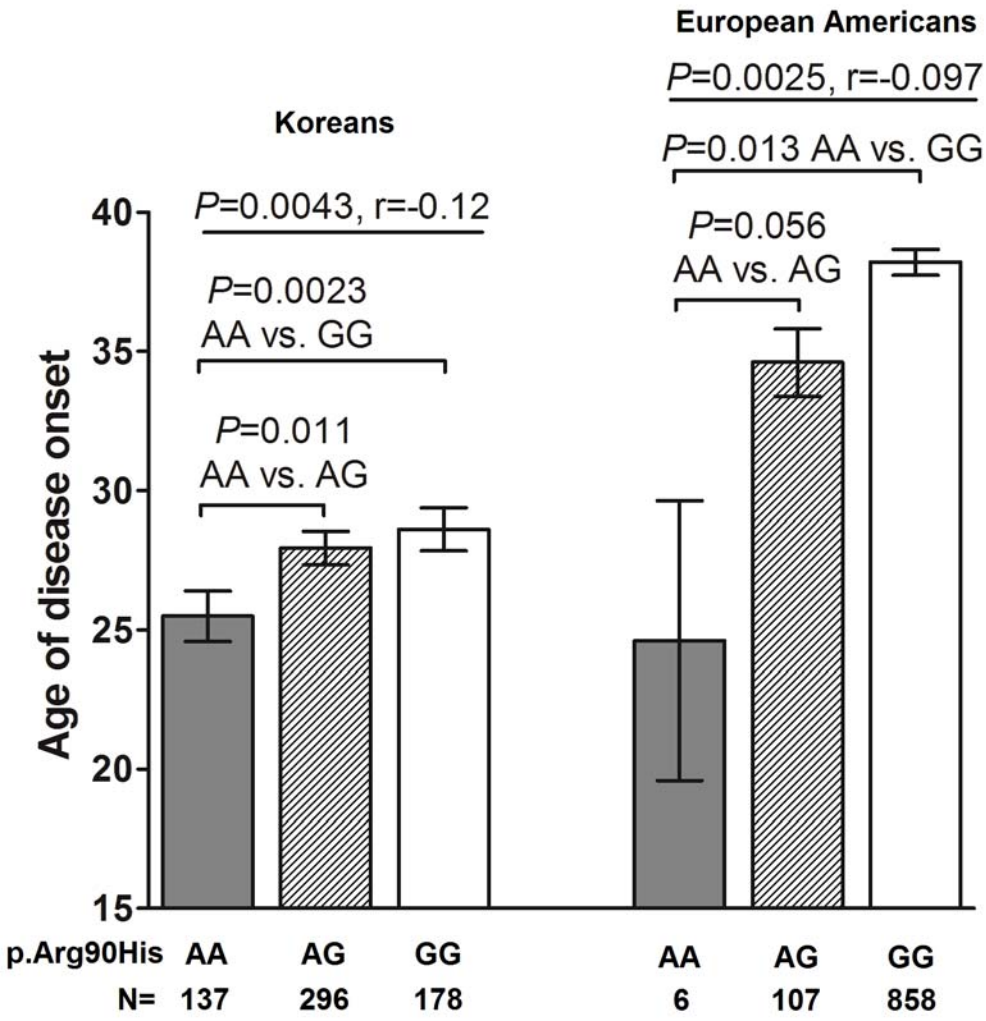
Supplementary Figures



Supplementary Fig 1. No association between rs117026326 genotypes and transcript levels of *NCF1*, *GTF2I* and *GTF2IRD1*. (a) Neighboring genes of rs117026326. Genes located within ±300 kb of rs117026326 include *GTF2IRD1*, *GTF2I*, *LOC101926943*, *NCF1*, *GTF2IRD2*, *STAG3L2*, *PMS2P5* and *GATSL2*. Of them, *NCF1*, which encodes the $p47^{phox}$ subunit of the NOX2 complex, is the most likely SLE-related gene. *GTF2I* encodes general transcription factor TFII-I; *GTF2IRD1* and *GTF2IRD2* encode structurally similar and potentially functionally overlapping TFII-I-like transcription factors; *LOC101926943* is an uncharacterized long noncoding RNA; *STAG3L2* and *PMS2P5* are both pseudogenes; and *GATSL2* encodes an arginine sensor for the mTORC1 pathway. (**b**) Association between rs117026326 genotypes and transcript levels of *NCF1*, *GTF2I* and *GTF2IRD1* in peripheral blood mononuclear cells (PBMCs) from patients with SLE and controls. Data were compared by Spearman correlation or Mann–Whitney test (two-tailed). Center lines and error bars represent means ± s.e.m.

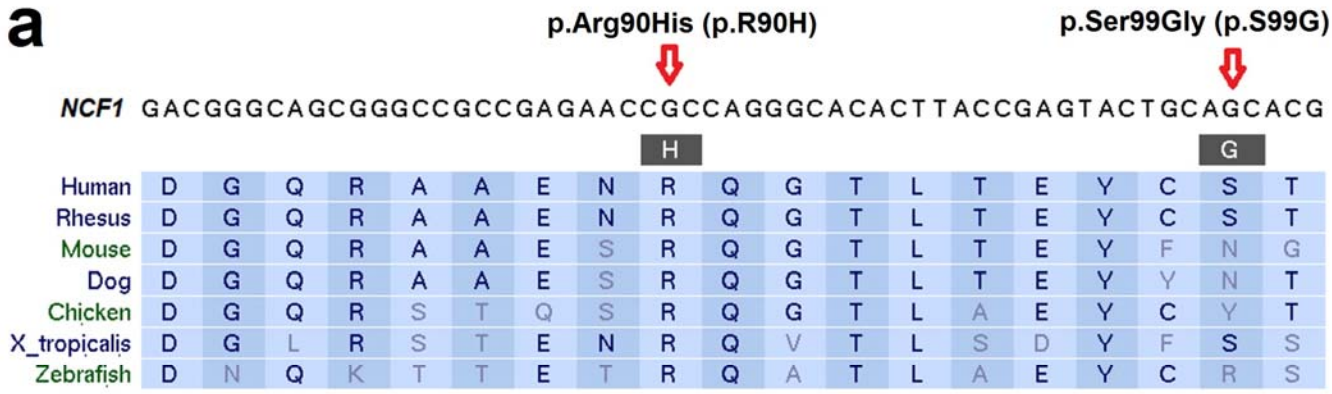


Supplementary Fig 2. PCR-amplification of *NCF1*-specific sequence. (a) *NCF1*-specific PCR primer binding sites. To exclude the influence of *NCF1B* and *NCF1C* and obtain correct genotypes of *NCF1* variants, we amplified *NCF1*-specific sequence by PCR. Two *NCF1*-specific loci were selected as PCR primer binding sites. One locus, targeted by PCR primers P1-R, P2-L and P2*-L as shown below in **b**, is a GTGT sequence at the beginning of exon 2 of *NCF1* (chr7:74,777,267–74,777,270), which is different from the GT deletion (ΔGT) in *NCF1B* and *NCF1C*. Another locus, targeted by PCR primer P3-L, is a T allele in intron 6 of *NCF1* (chr7:74,783,147), which is different from the G in *NCF1B* and *NCF1C*. (**b**) PCR amplification of *NCF1* for sequencing and SNP genotyping. The entire 15.5-kb region of *NCF1* was amplified by three PCR reactions (PCR products P1, P2 and P3) for Sanger sequencing. To genotype *NCF1* variants, we performed nested PCR and TaqMan assays, in which P2 (a larger PCR product containing p.Arg90His, p.Ser99Gly, intronic-1 and intronic-2) or P2* (a smaller PCR product containing p.Arg90His and p.Ser99Gly only) was obtained using *NCF1*-specific primer and then used as DNA template for TaqMan SNP genotyping assays.



Supplementary Fig 3. Significant association of p.Arg90His risk genotypes with early age of disease onset in Korean and European-American patients with SLE. Data were compared by Spearman correlation or Mann-Whitney test (two-tailed). Center lines and error bars represent means±s.e.m.

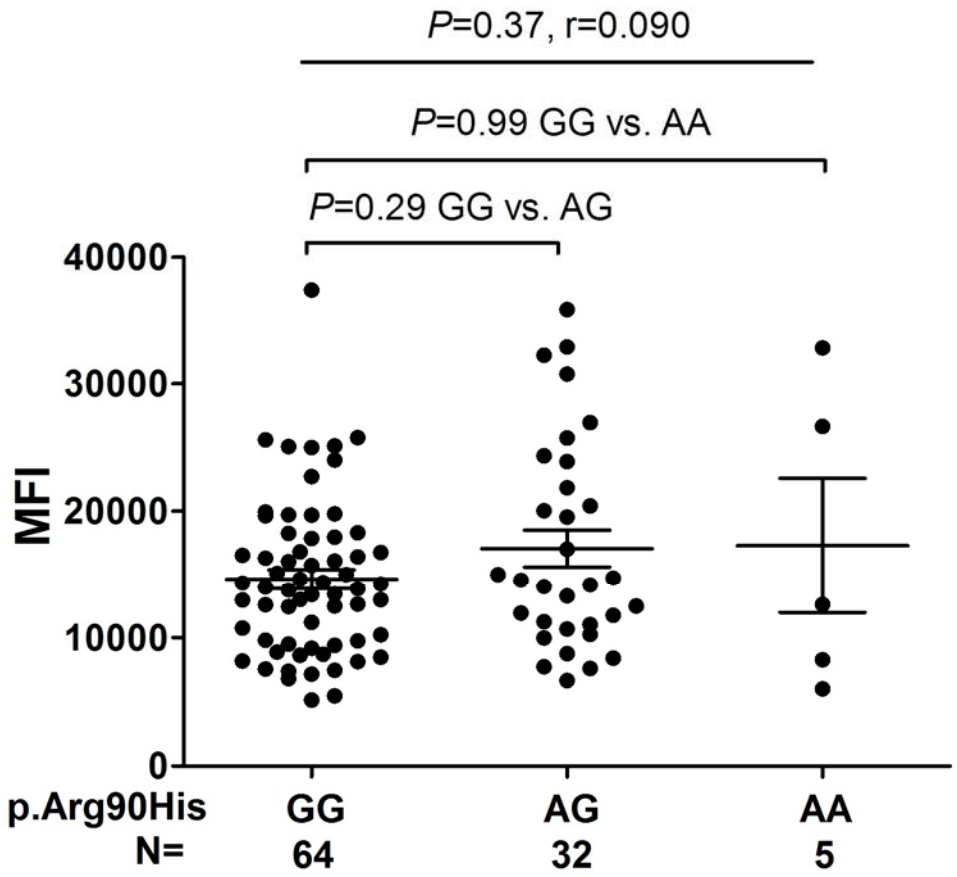
a



b

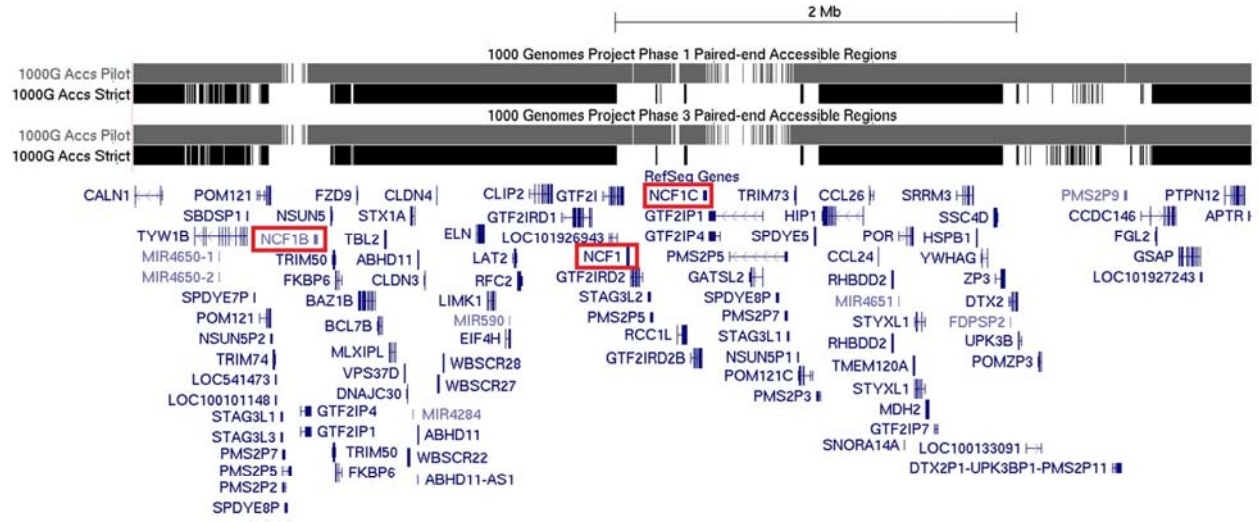
	Computational prediction					
	SIFT	Polyphen2	PANTHER	MutationTaster	MutationAssessor	FATHMM
p.Arg90His	Deleterious	Possibly damaging	91% prob. of deleterious	Disease causing	High functional impact	Damaging
p.Ser99Gly	Tolerated	Benign	35% prob. of deleterious	Polymorphism	Neutral functional impact	Tolerated

Supplementary Fig 4. Evolutionary conservation and computational prediction for functional impact of p.Arg90His and p.Ser99Gly. (a) Alignments of multiple vertebrate species at p.Arg90His and p.Ser99Gly. Arg90 is an evolutionarily conserved amino acid. This figure was adapted from the UCSC Genome Browser. (**b**) Assessment of the functional impact of p.Arg90His and p.Ser99Gly. The substitution of Arg90 with a histidine residue encoded by the SLE risk allele was predicted to be deleterious by softwares, including SIFT (Sorting Intolerant From Tolerant; http://sift.bii.a-star.edu.sg/), PolyPhen-2 (Polymorphism Phenotyping v2; http://genetics.bwh.harvard.edu/pph2/), PANTHER (Protein ANalysis THrough Evolutionary Relationships; http://www.pantherdb.org/), MutationTaster (http://www.mutationtaster.org/), MutationAssessor (http://mutationassessor.org/) and FATHMM (Functional Analysis through Hidden Markov Models; http://fathmm.biocompute.org.uk/).



Supplementary Fig 5. No association between p.Arg90His and ROS levels in neutrophils from healthy controls. Intracellular ROS levels were determined using fluorescent dye DCFH-DA and measured using flow cytometry Data were compared by Spearman correlation or Mann–Whitney test (two-tailed). Center lines and error bars represent means ± s.e.m.

a

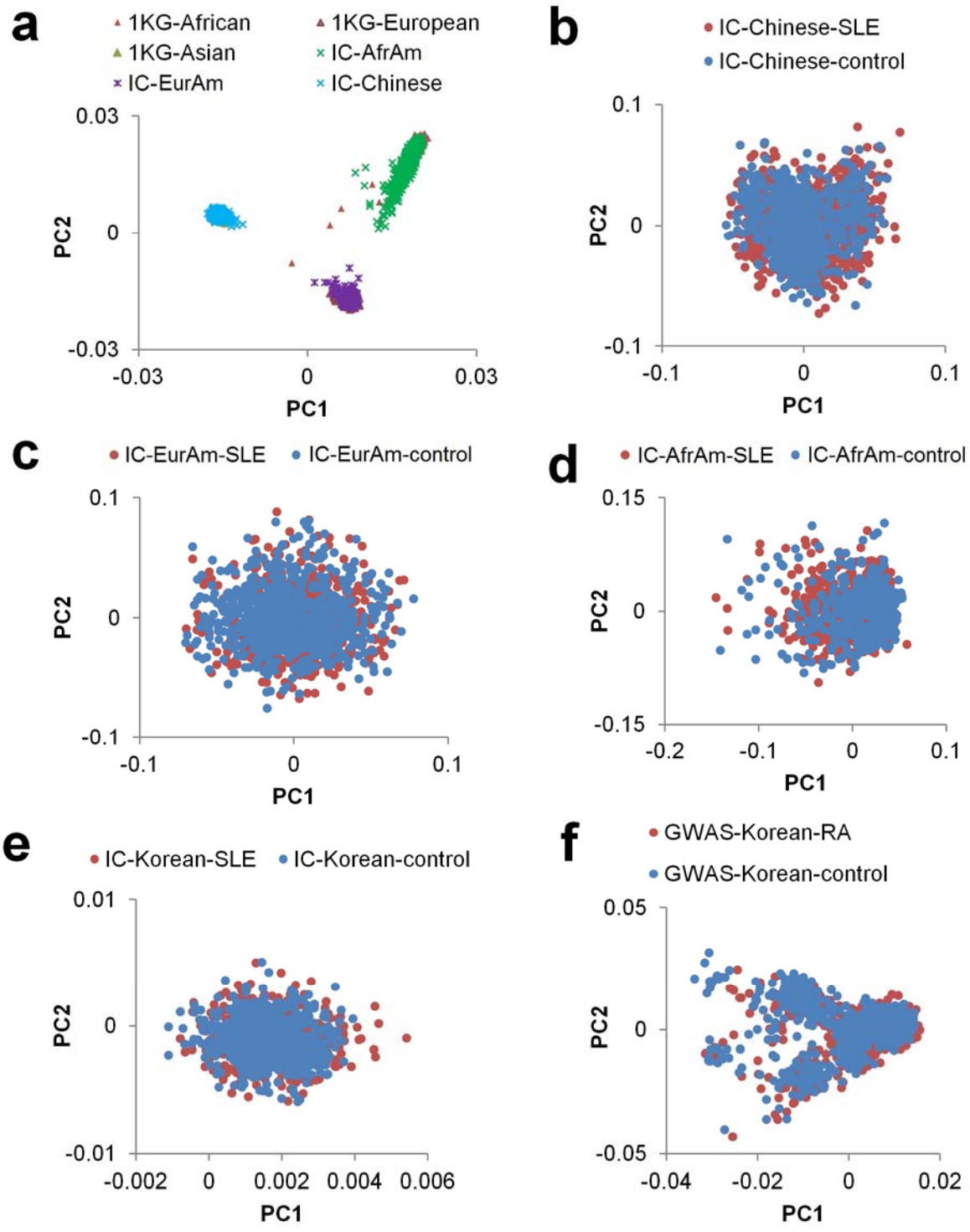


b

Phase	*NCF1* variants in 1000 Genomes	Asian		European		African	
		MAF	P_{HWE}	MAF	P_{HWE}	MAF	P_{HWE}
Phase1&3	rs140034807	1.5%	1.00	1.1%	1.00	2.6%	0.37
Phase1&3	rs138054188	2.6%	1.00	2.7%	1.00	4.8%	0.65
Phase3	rs377305075	0.0%	1.00	0.0%	1.00	0.5%	1.00
Phase3	rs377662255	0.0%	1.00	1.7%	1.00	0.0%	1.00
Phase3	rs587631188	0.0%	1.00	0.0%	1.00	1.1%	1.00
Phase3	rs587662134	0.0%	1.00	0.0%	1.00	4.5%	0.64
Phase3	rs200623471	4.1%	1.00	20.4%	0.68	11.4%	0.12
Phase3	rs368231459	19.2%	1.1E-18	25.8%	3.5E-24	12.5%	1.4E-23
Phase3	rs139225348	0.0%	1.00	1.0%	1.00	0.1%	1.00
Phase3	rs17295741 (p.Ser99Gly)	41.3%	3.3E-24	28.6%	6.7E-07	49.2%	7.6E-23
Phase3	rs373919021	0.1%	1.00	0.0%	1.00	9.5%	6.0E-05
Phase3	rs587757490	0.0%	1.00	0.0%	1.00	1.7%	0.16
Phase1&3	rs140969778	0.1%	1.00	0.0%	1.00	0.5%	1.00
Phase1&3	rs800978	3.9%	1.00	19.4%	0.48	15.1%	0.76
Phase3	rs587770703	10.2%	3.0E-12	33.3%	0.04	23.4%	9.6E-18
Phase3	rs199789198 (intronic-2)	3.5%	0.46	0.2%	1.00	0.1%	1.00
Phase3	rs800980	1.5%	1.00	15.8%	0.50	5.1%	0.40
Phase1&3	rs800981	4.1%	1.00	19.9%	0.328	10.0%	1.00
Phase3	rs62475426	4.8%	4.1E-04	8.0%	1.2E-05	43.9%	1.1E-07
Phase3	rs587683486	0.0%	1.00	0.7%	1.00	0.0%	1.00
Phase1&3	rs2528941	12.9%	1.3E-22	20.8%	5.6E-24	18.5%	3.1E-43
Phase3	rs587616286	0.1%	1.00	4.4%	0.62	2.6%	1.00
Phase3	rs587697744	0.0%	1.00	0.0%	1.00	2.2%	1.00
Phase3	rs587629774	0.0%	1.00	0.0%	1.00	4.6%	0.64
Phase3	rs587600267	0.8%	1.00	0.0%	1.00	0.0%	1.00
Phase3	rs587721998	0.0%	1.00	0.0%	1.00	3.4%	1.00
Phase3	rs372181124	0.6%	1.00	0.0%	1.00	0.0%	1.00
Phase3	rs369485834	4.1%	1.00	19.8%	0.40	8.8%	6.3E-03
Phase1&3	rs191081238	7.3%	0.10	4.9%	1.00	0.2%	1.00
Phase3	rs587619899	0.0%	1.00	0.0%	1.00	4.0%	1.00
Phase1&3	rs138406096	0.0%	1.00	0.0%	1.00	3.5%	1.00
Phase3	rs200877252	4.6%	0.02	7.0%	1.00	49.9%	0.94
Phase3	rs587640002	0.0%	1.00	0.0%	1.00	1.7%	1.00
Phase3	rs587619282	4.0%	1.00	18.8%	1.00	5.4%	0.25

Supplementary Fig 6. *NCF1* variants in the 1000 Genomes Project. (a) The 1000 Genomes Project inaccessible region at 7q11.23. The 'pilot' and 'strict' level of stringency in the 1000 Genomes Project are shown as gray and black bars on the top, respectively. *NCF1*, *NCF1B* and *NCF1C* are located in

regions that do not meet the ‘strict’ level of stringency in 1000 Genomes Project phases 1 and 3. This figure was adapted from the UCSC Genome Browser. (**b**) *NCF1* variants included in the 1000 Genomes Project. *NCF1* variants with MAF >0.5% in at least one ancestral group (*n* = 8 in phase 1; *n* = 34 in phase 3) are shown in this table. p.Arg90HisR90H (rs201802880) is not included in either phase 1 or 3. p.Ser99Gly (rs17295741) is included in phase 3, which however shows deviation from Hardy–Weinberg equilibrium (HWE). In addition, deviations from HWE are observed at the four other common *NCF1* SNPs (rs368231459, rs587770703, rs62475426 and rs2528941) included in phase 3, which indicates that 1000 Genomes Project data in the *NCF1* region are unreliable.



Supplementary Fig 7. Plots of the principal-component analysis (PCA). PCA of Chinese, European-American (EurAm) and African-American (AfrAm) samples genotyped by Immunochip (IC) along with reference samples from the 1000 Genomes Project. (**b**–**d**) PCA of Chinese, European-American and African-American subjects genotyped by IC. (**e**,**f**) PCA of Korean SLE cases, RA cases and healthy controls in the replication stage.