nature genetics

Functional variants in the B-cell gene *BANK1* are associated with systemic lupus erythematosus

Sergey V Kozyrev[1,15], Anna-Karin Abelson[1,15], Jerome Wojcik[2], Ammar Zaghlool[1], MV Prasad Linga Reddy[1], Elena Sanchez[3], Iva Gunnarsson[4], Elisabet Svenungsson[4], Gunnar Sturfelt[5], Andreas Jönsen[5], Lennart Truedsson[6], Bernardo A Pons-Estel[7], Torsten Witte[8], Sandra D'Alfonso[9], Nadia Barizzone[9], Maria Giovanna Danieli[10], Carmen Gutierrez[11], Ana Suarez[11], Peter Junker[12], Helle Laustrup[12], Maria Francisca González-Escribano[13], Javier Martin[3,14], Hadi Abderrahim[2] & Marta E Alarcón-Riquelme[1]

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

Systemic lupus erythematosus (SLE) is a prototypical autoimmune disease characterized by production of autoantibodies and complex genetic inheritance[1–3]. In a genome-wide scan using 85,042 SNPs, we identified an association between SLE and a nonsynonymous substitution (rs10516487, R61H) in the B-cell scaffold protein with ankyrin repeats gene, *BANK1*. We replicated the association in four independent case-control sets (combined $P = 3.7 \times 10^{-10}$; OR = 1.38). We analyzed *BANK1* cDNA and found two isoforms, one full-length and the other alternatively spliced and lacking exon 2 (Δ2), encoding a protein without a putative IP3R-binding domain. The transcripts were differentially expressed depending on a branch point–site SNP, rs17266594, in strong linkage disequilibrium (LD) with rs10516487. A third associated variant was found in the ankyrin domain (rs3733197, A383T). Our findings implicate *BANK1* as a susceptibility gene for SLE, with variants affecting regulatory sites and key functional domains. The disease-associated variants could contribute to sustained B cell–receptor signaling and B-cell hyperactivity characteristic of this disease.

We genotyped 279 Swedish individuals with SLE and 515 control individuals using the 100K Affymetrix SNP array. As our purpose was to identify non-MHC genes and, if possible, important functional polymorphisms involved in SLE pathogenesis, we carried out an analysis of the genomic location of the associated SNPs, focusing on nonsynonymous substitutions. Among all associated SNPs, one (rs10516487) led to a substitution of arginine to histidine at amino-acid position 61 (R61H) of the BANK1 protein (allelic association, $P = 6.4 \times 10^{-3}$; genotypic association, $P = 2.01 \times 10^{-2}$). This SNP was ranked 679th in the allelic and 2,148th in the genotypic tests across the genome scan, with estimated false-discovery rates of 71.1% and 77.5%, respectively[4]. Four other SNPs in *BANK1* also showed association (**Supplementary Table 1** online). Because of the described B cell–specific expression of *BANK1* and its potential role in B cell receptor–mediated activation, we pursued this gene[5,6].

To provide better SNP coverage and refine the association signal, we genotyped 30 SNPs spanning the 284-kb *BANK1* genomic region (including the scan SNPs) in the Swedish SLE case and control samples. Two SNPs were not polymorphic, and nine SNPs were associated (**Table 1** bold). All associated SNPs were located between introns 1 and 7 (**Table 1**, **Supplementary Table 2** and **Supplementary Fig. 1** online).

Next, we carried out a detailed analysis of *BANK1* expression and structure. We observed that *BANK1* is indeed primarily expressed in $CD19^+$ B cells, with very low expression in other cell populations (**Fig. 1a**). To clone *BANK1* for functional analysis, we amplified full-length cDNA using distal primers. Of note, we detected two main bands following gel electrophoresis of the PCR products (**Fig. 1b**). *BANK1* is known to have two full-length alternative isoforms containing exon 1A or exon 1B[5]. Through subsequent cloning and sequencing, we identified a previously unknown isoform with an in-frame deletion of the entire exon 2 (Δ2 isoform). We analyzed cDNA from 83 healthy individuals and 30 individuals with SLE and found that this isoform was present in every sample, indicating that it is constitutively spliced. Moreover, we detected this isoform in cDNA from chimpanzee and mouse spleen (data not shown), suggesting that it is conserved across species.

[1]Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala 75185, Sweden. [2]Merck Serono, Chemin des Mines, Geneva 12202, Switzerland. [3]Instituto de Biomedicina López-Neyra, Granada 18100, Spain. [4]Department of Medicine, Unit of Rheumatology, Karolinska University Hospital, Solna 17176, Sweden. [5]Department of Rheumatology, University of Lund, Lund 22200, Sweden. [6]Department of Clinical Immunology and Microbiology, University of Lund, Lund 22200, Sweden. [7]Sanatorio Parque, Rosario 2000, Argentina. [8]Medical School Hannover, Hannover 30625, Germany. [9]Department of Medical Sciences and Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara 13100, Italy. [10]Dipartimento di Scienze Mediche e Chirurgiche, Università Politecnica delle Marche, Ancona 60121, Italy. [11]Department of Functional Biology, Hospital Universitario Central de Asturias, Universidad de Oviedo, Oviedo 33006, Spain. [12]Department of Internal Medicine C, Section of Rheumatology, Odense University Hospital, Odense 5000, Denmark. [13]Departamento de Inmunología, Hospital Virgen del Rocío, Sevilla 41013, Spain. [14]Consejo Superior de Investigaciones Científicas (CSIC), Grenada 18100, Spain. [15]These authors contributed equally to this work. Correspondence should be addressed to M.E.A.-R. (marta.alarcon@genpat.uu.se).

Received 4 September 2007; accepted 12 December 2007; published online 20 January 2008; corrected after print 27 March 2008; doi:10.1038/ng.79

Table 1 Association analysis of *BANK1* SNPs in Swedish SLE cases and controls

SNP	Associated allele	χ^2	*P*
rs7675129	T	0.155	0.6933
rs11726012	G	0.361	0.5479
rs11097755	C	1.433	0.2313
rs4522865	A	6.757	**0.0093**
rs4496585	A	2.618	0.1057
rs4572885	T	5.113	**0.0238**
rs10516487	G	8.091	**0.0044**
rs10516486	C	9.62	**0.0019**
rs17200824	A	4.265	**0.0389**
rs6849308	C	7.268	**0.007**
rs10516482	C	8.625	**0.0033**
rs10516483	C	11.437	**0.0007**
rs10516484	A	0.595	0.4404
rs4493533	C	3.184	0.0744
rs3733197	G	0.692	0.4054
rs2631271	G	6.356	**0.0117**
rs2850390	C	1.225	0.2684
rs2631265	C	0.016	0.8997
rs2631267	G	0.1	0.7524
rs2631268	T	1.446	0.2292
rs10516491	C	2.424	0.1195
rs1872701	G	1.842	0.1747
rs2850393	T	0.061	0.8053
rs2850396	C	0.682	0.4088
rs10516490	G	0.329	0.5665
rs10516489	T	0.338	0.5609
rs10516488	G	0.561	0.4538
rs1395306	T	1.97	0.1604

We carried out quantitative analysis of isoform expression in peripheral blood mononuclear cells. As exon 1B transcript was present at very low concentrations (data not shown), we continued the analysis, measuring common (exon 1A and exon 1B) full-length isoform concentrations. We noticed that the ratio of the full-length isoform to Δ2 (FL/Δ2) was not constant, which would be expected if Δ2 were equally expressed regardless of the genotypes of the samples. On the contrary, samples could be divided into groups according to the FL/Δ2 isoform ratio. After closely examining the genomic sequences surrounding exon 2, where putative signals affecting splicing could be located, we identified one SNP, rs17266594, located in the branch-point site. When we re-grouped the expression data, we observed a clear difference between genotypes (**Fig. 1c**). Individuals homozygous for the T allele and having the classical structure of the branch-point site[7] (YNYTGAYYN) showed higher expression of the full-length isoform; this expression was significantly suppressed (up to 40%) in homozygotes for the minor allele C, with concomitant upregulation of the Δ2 isoform expression. Total *BANK1* expression was not significantly affected by the SNP (**Fig. 1d**).

To determine whether other polymorphisms might contribute to the alternative splicing of exon 2, we sequenced the proximal promoter regions, exon 1A, exon 1B and exon 2 and 500 bp upstream and downstream of these exons, in 24 individuals with SLE and 8 controls. However, we found no previously unidentified SNPs in these regions that could be functional. Next, we identified five nonsynonymous substitutions in *BANK1* from the SNP databases. Although most were nonpolymorphic in our samples, we identified one polymorphic SNP, rs3733197, causing an alanine to threonine substitution at amino acid position 383 (A383T) in exon 7, which encodes the ankyrin repeat-like motif (**Supplementary Table 3** online).

To extend our association analyses, we genotyped rs10516487 (R61H), rs17266594 (branch-point variant) and rs3733197 (A383T) in four additional sets of SLE cases and controls (**Table 2**). We corroborated the genetic association with SLE for all three SNPs, although there were differences between individual populations. Using homogeneity and combinability tests according to the Breslow-Day method, we carried out a meta-analysis comprising 3,971 individuals. We then used the Mantel-Haenszel test to calculate pooled odds ratios of 1.38 ($P = 3.74 \times 10^{-10}$), 1.42 ($P = 4.74 \times 10^{-11}$) and 1.23

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

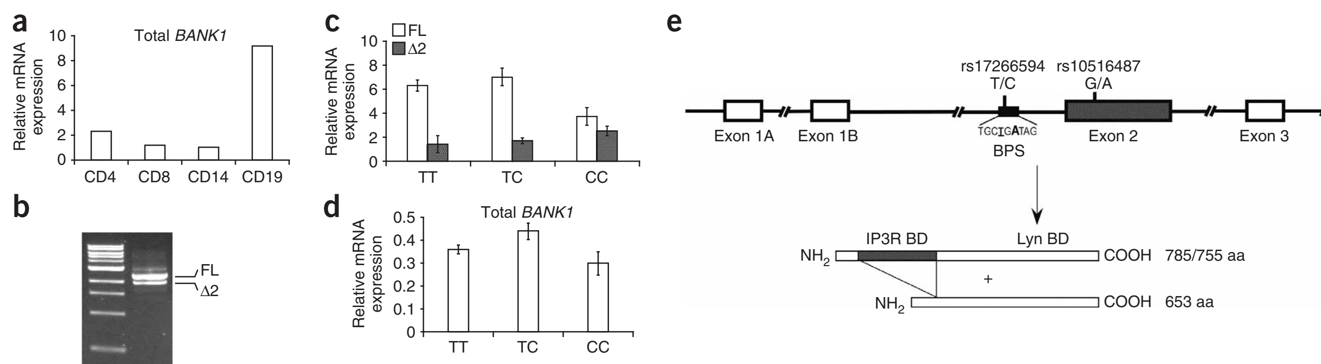

Figure 1 Correlation of rs17266594 genotypes with differences in FL/Δ2 isoform ratio of *BANK1*. (**a**) Expression of total *BANK1* mRNA in Clontech human blood fractions: CD4+ and CD8+ T cells, CD14+ myeloid cells and CD19+ B cells. (**b**) RT-PCR of the coding sequence of *BANK1* amplified from total human spleen cDNA shows two bands on a gel. Ladder (1 kb, New England Biolabs) is shown on the left. We confirmed the identity of both bands (2.3-kb upper band and 1.9-kb smaller band) by sequencing. (**c**) Relative mRNA expression of the full-length and Δ2 isoforms, as determined by quantitative real-time RT-PCR on total RNA purified from human PBMCs of healthy controls and cases. Data represent mean ± s.d. We analyzed 39 individuals with TT for the branch point–site SNP, 34 with TC and 10 with CC. Full-length transcript: TT versus CC, $P = 0.0004$ (Student's *t*-test); Δ2 transcript: TT versus CC, $P = 0.0088$. (**d**) Total *BANK1* expression did not correlate with genotypes of rs17266594. TT versus CC, $P = 0.229$. (**e**) Schematic structure of the 5′ end of *BANK1*. rs17266594, located in the branch point site of intron 1, alters splicing efficiency of the full-length and Δ2 transcripts. rs10516487 results in nonsynonymous substitution R61H. Alternative splicing gives rise to two isoforms, full-length and Δ2 with an in-frame deletion of entire exon 2. The short protein isoform lacks the putative domain for IP3R binding. IP3R BD, inositol 1,4,5-triphosphate receptor binding domain; Lyn BD, tyrosine kinase Lyn binding domain.

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

Table 2 Genotypic and allelic association of rs10516487 (R61H), rs17266594 and rs3733197 in five sets of SLE cases and controls and joint analysis with Mantel-Haenszel test

	Population	n	GG	GA	AA	χ^2	P	Odds ratio (CI)[a]	Allele G	Allele A	P	Odds ratio (CI)[b]
rs10516487	Scandinavia	Cases (536)	309 (57.6%)	200 (37.3%)	27 (5.0%)	11.7874	0.0028	GG: 2.12 (1.29–3.47)	818 (76.3%)	254 (23.7%)	7.3×10^{-4}	1.39 (1.14–1.68)
		Controls (565)	276 (48.8%)	238 (42.1%)	51 (9.0%)			GA: 1.59 (0.96–2.63)	790 (69.9%)	340 (30.1%)		
	Argentina	Cases (255)	164 (64.3%)	75 (29.4%)	16 (6.3%)	3.8013	0.1495	GG: 1.41 (0.73–2.72)	403 (79.0%)	107 (21.0%)	0.0564	1.31 (0.98–1.74)
		Controls (337)	190 (56.4%)	121 (35.9%)	26 (7.7%)			GA: 1.01 (0.51–2.00)	499 (74.3%)	173 (25.7%)		
	Germany	Cases (312)	181 (58.0%)	118 (37.8%)	13 (4.2%)	11.8503	0.0027	GG: 2.60 (1.32–5.14)	480 (76.9%)	144 (23.1%)	8.13×10^{-4}	1.52 (1.18–1.95)
		Controls (360)	166 (46.1%)	163 (45.3%)	31 (8.6%)			GA: 1.73 (0.87–3.44)	495 (68.8%)	225 (31.2%)		
	Italy	Cases (279)	166 (59.5%)	100 (35.8%)	13 (4.7%)	7.5139	0.0234	GG: 2.49 (1.22–5.09)	432 (77.4%)	126 (22.6%)	0.0078	1.46 (1.09–1.94)
		Controls (245)	123 (50.2%)	98 (40.0%)	24 (9.8)			GA: 1.88 (0.91–3.91)	344 (70.2%)	146 (29.8%)		
	Spain	Cases (702)	414 (59.0%)	243 (34.6%)	45 (6.4%)	11.3579	0.0034	GG: 1.26 (0.77–2.06)	1,071 (76.3%)	333 (23.7%)	0.0065	1.30 (1.07–1.58)
		Controls (446)	219 (49.1%)	197 (44.2%)	30 (6.7%)			GA: 0.82 (0.50–1.35)	635 (71.2%)	257 (28.8%)		
	Pooled	Cases (2,003)	1,187 (59.3%)	706 (35.2%)	110 (5.5%)				3,080 (76.9%)	926 (23.1%)	3.74×10^{-10}[c]	1.38 (1.25–1.53)
		Controls (1,968)	974 (49.9%)	817 (41.8%)	162 (8.3%)				2,763 (70.8%)	1,141 (29.2%)		
	Population		TT	CT	CC	χ^2	P	Odds ratio (CI)[a]	Allele T	Allele C	P	Odds ratio (CI)
rs17266594	Scandinavia	Cases (511)	296 (57.9%)	189 (37.0%)	26 (5.1%)	9.4399	0.0089	TT: 2.17 (1.28–3.66)	781 (76.4%)	241 (23.6%)	0.0036	1.36 (1.10–1.68)
		Controls (416)	210 (50.5%)	166 (39.9%)	40 (9.6%)			CT: 1.75 (1.03–2.99)	586 (70.4%)	246 (29.6%)		
	Argentina	Cases (274)	188 (68.6%)	77 (28.1%)	9 (3.3%)	14.1697	8.38×10^{-4}	TT: 3.26 (1.51–7.06)	453 (82.7%)	95 (17.3%)	1.06×10^{-4}	1.73 (1.30–2.31)
		Controls (346)	192 (55.5%)	124 (35.8%)	30 (8.7%)			CT: 2.07 (0.93–4.59)	508 (73.4%)	184 (26.6%)		
	Germany	Cases (241)	132 (54.8%)	98 (40.7%)	11 (4.6%)	7.7164	0.0211	TT: 2.46 (1.19–5.09)	362 (75.1%)	120 (24.9%)	0.0080	1.43 (1.09–1.87)
		Controls (335)	151 (45.1%)	153 (45.7%)	31 (9.3%)			CT: 1.81 (0.87–3.76)	455 (67.9%)	215 (32.1%)		
	Italy	Cases (231)	130 (56.3%)	87 (37.7%)	14 (6.1%)	10.1706	0.0062	TT: 2.42 (1.19–4.93)	347 (75.1%)	115 (24.9%)	0.0016	1.59 (1.18–2.14)
		Controls (219)	92 (42.0%)	103 (47.0%)	24 (11.0%)			CT: 1.45 (0.71–2.97)	287 (65.5%)	151 (34.5%)		
	Spain	Cases (678)	404 (59.6%)	231 (34.1%)	43 (6.3%)	14.8617	5.93×10^{-4}	TT: 1.04 (0.62–1.76)	1,039 (76.6%)	317 (23.4%)	0.010	1.29 (1.06–1.56)
		Controls (458)	225 (49.1%)	208 (45.4%)	25 (5.5%)			CT: 0.65 (0.38–1.09)	658 (71.8%)	258 (28.2%)		
	Pooled	Cases (1,856)	1,102 (59.4%)	655 (35.3%)	99 (5.3%)				2,859 (77.0%)	853 (23.0%)	4.74×10^{-11}	1.42 (1.28–1.58)
		Controls (1,774)	870 (49.0%)	754 (42.5%)	150 (8.5%)				2,494 (70.3%)	1,054 (29.7%)		
	Population		GG	GA	AA	χ^2	P	Odds ratio (CI)[a]	Allele G	Allele A	P	Odds ratio (CI)
rs3733197	Scandinavia	Cases (419)	167 (39.9%)	192 (45.8%)	60 (14.3%)	1.2365	0.5389	GG: 1.04 (0.69–1.58)	526 (62.8%)	312 (37.2%)	0.5832	1.06 (0.87–1.29)
		Controls (444)	163 (36.7%)	220 (49.6%)	61 (13.7%)			GA: 0.89 (0.59–1.33)	546 (61.5%)	342 (38.5%)		
	Argentina	Cases (287)	177 (61.7%)	97 (33.8%)	13 (4.5%)	9.6496	0.0080	GG: 2.36 (1.20–4.66)	451 (78.6%)	123 (21.4%)	0.0018	1.15 (0.95–1.40)
		Controls (363)	184 (50.7%)	147 (40.5%)	32 (8.8%)			GA: 1.62 (0.81–3.25)	515 (70.9%)	211 (29.1%)		
	Germany	Cases (272)	128 (47.1%)	112 (41.2%)	32 (11.8%)	4.1431	0.1260	GG: 1.65 (1.01–2.69)	368 (67.6%)	176 (32.4%)	0.0382	1.28 (1.00–1.63)
		Controls (362)	148 (40.9%)	153 (42.3%)	61 (16.9%)			GA: 1.40 (0.85–2.28)	449 (62.0%)	275 (38.0%)		
	Italy	Cases (253)	131 (51.8%)	102 (40.3%)	20 (7.9%)	8.2595	0.0161	GG: 1.74 (0.92–3.29)	364 (71.9%)	142 (28.1%)	0.0097	1.42 (1.08–1.87)
		Controls (251)	98 (39.0%)	127 (50.6%)	26 (10.4%)			GA: 1.04 (0.55–1.98)	323 (64.3%)	179 (35.7%)		
	Spain	Cases (588)	307 (52.2%)	234 (39.8%)	47 (8.0%)	3.4580	0.1775	GG: 1.14 (0.72–1.82)	977 (72.1%)	379 (27.9%)	0.1474	1.50 (1.15–1.96)
		Controls (455)	212 (46.6%)	206 (45.3%)	37 (8.1%)			GA: 0.89 (0.56–1.43)	630 (69.2%)	280 (30.8%)		
	Pooled	Cases (1,819)	910 (50.0%)	737 (40.5%)	172 (9.5%)				2,686 (70.4%)	1,132 (29.6%)	4.67×10^{-5}	1.23 (1.11–1.36)
		Controls (1,875)	805 (42.9%)	853 (45.5%)	217 (11.6%)				2,463 (65.7%)	1,287 (34.3%)		

[a]Genotypic odds ratio calculated using homozygosity for the protective allele as reference with OR = 1. [b]Calculated using the Robins, Breslow and Greenland method. [c]Calculated using the Mantel-Haenszel χ^2 with fixed effects.

Table 3 Effect sizes of individual SNPs and 2- and 3-SNP haplotypes of *BANK1*[a]

SNP or haplotype	Allele or haplotype	Frequency	OR	95% CI	Effect
rs17266594 (branch point)	T	0.738	1.18	1.09–1.23	Risk
	C	0.262	0.70	0.63–0.78	Protection
rs10516487 (R61H)	G	0.739	1.16	1.07–1.26	Risk
	A	0.261	0.72	0.66–0.80	Protection
rs3733197 (A383T)	G	0.680	1.13	1.04–1.23	Risk
	A	0.320	0.80	0.73–0.89	Protection
2-SNP haplotype[b]	TG	0.744	1.16	1.07–1.25	Risk
	CA	0.256	0.70	0.63–0.77	Protection
3-SNP haplotype	TGG	0.636	1.16	1.06–1.27	Risk
	CAA	0.211	0.69	0.62–0.77	Protection
	TGA	0.108	0.98	0.84–1.14	Neutral
	CAG	0.045	0.73	0.54–0.92	Protection

[a]Effect sizes were calculated using WHAP, and ORs were estimated using R language.
[b]2-SNP = rs17266594 + rs10516487; 3-SNPs are in the order rs17266594, rs10516487 and rs3733197.

(P = 4.67 × 10^{-5}) for rs10516485, rs17266594 and rs3733197, respectively, for the allelic association, supported by genotypic association (**Table 2**).

rs17266594 and rs10516487 are separated by 153 bp and are in strong LD (D′ = 0.95; r^2 = 0.90 calculated for all sets jointly; **Supplementary Fig. 2** online). rs3733197 is 88 kb away from rs10516487 (D′ = 0.72; r^2 = 0.39) and rs17266594 (D′ = 0.68; r^2 = 0.27), and could segregate with a risk haplotype in some individuals (**Supplementary Figs. 2** and **3** online). To better define the relative contribution of each SNP, we carried out conditional logistic regression analyses using the three SNPs. We found that none of the SNPs is independent of the others, as a result of the LD between them (colineality in the multiple logistic regression analysis). Through haplotype-based logistic regression analysis using WHAP[8], we did not find any differences in the effect sizes (OR) of the individual SNP alleles or the 2- or 3-SNP haplotypes (**Table 3**). Thus, linkage disequilibrium, haplotype and conditional regression analyses suggested that all three SNPs, either individually or as haplotypes, confer susceptibility for SLE.

BANK1 is a B-cell adaptor protein[9,10]. The two full-length isoforms of 785 and 755 amino acids differ by 30 amino acids at the N terminus, encoded by the alternative exon 1A (**Fig. 1e**), and contain ankyrin repeat motifs and coiled-coil regions, structures very similar to other adaptor proteins[11]. B-cell activation through the B-cell receptor leads to tyrosine phosphorylation of BANK1, which in turn promotes its association with the tyrosine kinase Lyn and the calcium channel IP3R, facilitating phosphorylation and activation of IP3R by Lyn and release of Ca^{2+} from endoplasmic reticulum stores[5,12]. IP3R associates with the N-terminal domain of BANK1 encoded by exon 2, whereas Lyn interacts with the C-terminal portion[5]. Our own analysis predicts a pleckstrin homology domain in the N terminus, which could also participate in phosphatidylinositol-mediated signaling. rs10516487 lies within the region essential for binding of IP3R. We speculate that R61, being highly protonated under conditions of physiological pH, could potentially alter the affinity of BANK1 for IP3R, favoring stronger binding, although this has yet to be tested.

rs17266594 may affect the relative splicing efficiency, but not splicing *per se*, of the full-length and Δ2 isoforms of *BANK1*. Mutations affecting the thymidine of the branch point consensus sequence and altering splicing efficiency have been previously described[13]. Through more efficient splicing of a full-length transcript containing the arginine residue in the IP3R-binding domain, a more ‘active’ protein would be expected in individuals at risk. On the contrary, given that the Δ2 isoform lacks the entire exon 2 and thus the IP3R-binding and PH domains (**Fig. 1e**), it possibly functions as a dominant-negative or rather, a dose-dependent isoform attenuating BANK1-mediated signaling. This is supported by the observation of a strong protective genetic effect in individuals with the CC and CT genotypes of rs17266594 (CC: OR = 0.52, 95% CI = 0.0–0.67; CT: OR = 0.68, 95% CI = 0.59–0.78) that show increased concentrations of the Δ2 isoform relative to the full-length isoform (**Fig. 1c**). Experimental evidence for a dominant-negative effect of the Δ2 isoform is needed to validate this proposed mechanism.

The importance of mutations in ankyrin motifs for interactions with IP3R was recently highlighted by a discovery linking single amino acid substitutions in the adaptor protein ankyrin-B with cardiac arrhythmia and sudden cardiac death[14]. Although the A383 variant is associated with SLE, the minor allele 383T of rs3733197 might create a site for threonine kinases[15].

B cells are the primary cell type affected in SLE. Novel therapies are aimed at depleting hyperactivated B cells that may function as autoantibody-producing cells and as important regulators of innate and adaptive immune responses through antigen presentation and cytokine-mediated signaling[16]. Functional and expression abnormalities of signaling molecules in B cells have been described in individuals with lupus. Of note, Lyn, a binding partner of BANK1, is of key importance in both human and mouse lupus disease models[17–22].

Increased binding of BANK1 to downstream effector proteins may lead to a steady state marked by B-cell hyperresponsiveness or deregulated B-cell activation. The precise role of BANK1 in B cell receptor–mediated signaling remains unclear, as two reports published to date contain conflicting data regarding the stimulatory or inhibitory role of BANK1 on B-cell activation[5,6]. Given the previously unreported existence of the alternative splicing of exon 2, we can speculate that the negative role for BANK1 assigned from the knockout model was a result of, in part, the residual expression of the Δ2 isoform, as this exon was targeted by the knockout construct[6]. Further experiments are required to fully understand if, and how, *BANK1* polymorphisms lead to B-cell hyperactivity, breakage of B-cell tolerance and production of autoantibodies, which are the principal hallmarks of SLE.

METHODS

Clinical samples. In our initial scan, we genotyped 279 Swedish SLE cases and 515 controls using the Affymetrix 100K array. Of these, 279 cases and 352 controls were available for the additional genotyping of *BANK1* SNPs (**Table 1**). For the functional polymorphisms, we genotyped an additional 185 Swedish SLE cases; 465 of the controls were available for genotyping of rs17266594 and rs3733197. For the final Mantel-Haenszel analysis and OR estimation, we added 84 Danish SLE cases to the Swedish cases, comprising the Scandinavian set (**Table 2**). The replication sets included 384 North German SLE cases and 374 controls, 288 Argentine SLE cases and 372 controls, and 286 Italian SLE cases and 252 controls. The Spanish cohort included 799 SLE cases and 542 controls from several regions in Spain. Of these, we genotyped 707 SLE cases and 469 controls for rs10516487 and rs3733197, and 678 SLE cases and 457 controls for rs17266594, as DNA from a number of controls was not available. The German, Spanish and Argentine SLE cases have all been previously described[23]. The Italian cases are a multicenter collection of affected individuals and their matched controls from Rome, Siena, Milan and Naples (North- and Mid-Italy). All cases fulfill the 1982 American College of Rheumatology (ACR) criteria for

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

the classification of SLE[24]. All participating subjects provided informed consent for this study. The study was approved by the various institutional review boards and ethical committees at each of the participating locations.

Genotyping. We carried out genotyping using the 100K Affymetrix array according to the manufacturer's instructions. We carried out fine-mapping and replication for SNPs rs10516487, rs17266594 and rs3733197 using TaqMan SNP genotyping assays (Applied Biosystems). The Affymetrix genotyping and fine-mapping were done at Serono Genetics Institute (now MerckSerono SA). The functional polymorphism replications were done at Uppsala University. For verification, 106 samples were genotyped twice, showing 100% concordance. Genotyping success rate for all the samples was over 92%.

Statistical analysis. For the 100K Affymetrix whole-genome scan analysis, we applied the following pre-processing filters. Specifically, SNPs were discarded if: (i) the proportion of missing genotypes was higher than 5%, (ii) the relative minor allele frequency was lower than 1%, or (iii) the probability that the observed genotype distribution results from sampling a SNP which follows the Hardy-Weinberg equilibrium was lower than 0.02. After filtering, we used data from 85,042 SNPs. We retained only SNPs from autosomal chromosomes for the sake of homogeneity between male and female subjects. SNP sequences were mapped onto NCBI 36 human genome assembly, and SNPs with multiple localizations were discarded. For each remaining SNP, we calculated genotypic and allelic frequencies in cases and controls and computed the corresponding probability values using exact (non-asymptotic) and unbiased algorithms[25]. Detailed results from the scan will be published elsewhere. The false-discovery rate (FDR) was then estimated using a method previously described[4].

For fine-mapping analyses, we estimated genetic association, haplotype estimation, LD and r^2 using Haploview (v4.0RC2). The Breslow-Day test of combinability and the Mantel-Haenszel test were carried out using the StatsDirect software (v2.4.6). As the Breslow-Day test showed combinability of the strata, the Mantel-Haenszel test for fixed effects was used in the analysis. Haplotypes were constructed using the PHASE software (v2.1)[26,27]. Genotypic odds ratios were calculated using the Unphased software (v3.0.9)[28].

We carried out logistic regression analysis and conditional multiple logistic regression analysis using R language glm routines. Haplotype-based logistic regression analysis was done using WHAP[8]. Coefficients were estimated with WHAP, and ORs and confidence intervals (**Table 3**) were calculated using R language.

Sequencing. DNA fragments for sequencing were amplified with the corresponding primers (**Supplementary Table 3**), purified from agarose gel with QIAquick gel extraction kit (Qiagen) and sequenced using BigDye Terminator 3.1 (Applied Biosystems) at the Uppsala Genome Center.

RNA purification and *BANK1* expression analysis. Total RNA was purified with TRIZOL Reagent (Invitrogen) from peripheral blood mononuclear cells (PBMCs) obtained with agreed consent from healthy donors and SLE cases. 2 μg of RNA were reverse-transcribed with 2 U of MultiScribe transcriptase in PCR buffer II containing 5 mM $MgCl_2$, 1 mM dNTPs, 0.4 U of RNase inhibitor and 5 μM oligo-dT. All reagents were purchased from Applied Biosystems. cDNA synthesis was done at 42 °C for 80 min, and then the reaction was terminated at 95 °C for 5 min. All cDNA samples were diluted to 15 ng/μl.

BANK1 expression was determined by real-time PCR on an ABI PRISM 7700 Sequence Detector (Applied Biosystems) with SDS 1.9.1 software. Total *BANK1*, both alternative full-length isoforms and the Δ2 isoform were quantified with SYBR Green and relevant primers (**Supplementary Table 3**). We carried out initial denaturation at 95 °C for 5 min followed by 45 cycles of PCR (95 °C for 15 s, 62 °C for 15 s and 72 °C for 30 s). PCR buffer provided with enzyme was supplemented with 3 mM $MgCl_2$, 200 μM of each of dNTPs, primers, SYBR Green (Molecular Probes), 15 ng of cDNA and 0.5 U of Platinum Taq polymerase (Invitrogen). Expression levels were normalized to the levels of TBP in the same samples amplified with commercial reagents (Applied Biosystems). All experiments were run in triplicate. Independent cDNA synthesis was carried out twice.

Expression levels for total *BANK1*, both full-length isoforms and Δ2 isoform in separated blood cell populations ($CD4^+$, $CD8^+$, $CD14^+$ and $CD19^+$ cells) were determined using Human Blood Fractions MTC Panel (Clontech). Statistical calculations were performed with available online GraphPad Software using two-tailed *t*-test.

Cloning of human, mouse and chimpanzee Δ2 isoform. Purification of total RNA from mouse spleen and cDNA synthesis were conducted as described above for the human PBMCs. Total RNA from chimpanzee (*Pan troglodytes*) spleen was provided by T. Bergström and L. Cavelier (Uppsala University). The human gene was amplified from Human Spleen BD Marathon-Ready cDNA (Clontech). After initial denaturation at 95 °C for 5 min, 35 cycles (95 °C for 20 s, 60 °C for 15 s and 72 °C for 2 min 30 s) were performed in PCR buffer containing 2 mM $MgSO_4$, 200 μM of each of dNTPs, 0.4 μM of each of the corresponding primers (**Supplementary Table 3**) and 0.5 U of Platinum Taq-High Fidelity enzyme (Invitrogen). Chimpanzee cDNA was amplified with human-specific primers. PCR products were purified from agarose gel and cloned in pCR 4-TOPO vector (Invitrogen) according to the manufacturer's instructions. Plasmid DNA from positive clones was purified with QIAprep Spin Miniprep kit (Qiagen) and verified by sequencing.

Accession codes. GenBank: full-length isoforms of *BANK1* containing exon 1A or exon 1B, NM_017935 and AB063170, respectively. Δ2 transcript sequences have been deposited with the following accession codes: EU051376, human; EU051377, chimpanzee and EU051378, mouse.

URLs. Haploview, http://www.broad.mit.edu/mpg/haploview/; GraphPad Software, http://www.graphpad.com; protein analysis, http://www.ebi.ac.uk/saps/, http://smart.embl-heidelberg.de/, http://ca.expasy.org/prosite/, and http://www.cbs.dtu.dk/services/NetPhos/; WHAP, http://pngu.mgh.harvard.edu/~purcell/whap/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS
The authors would like to express their gratitude to S. Lewén for help with purification of PBMCs and total RNA, H. Yin for help with the preparation of DNA samples and to the members of the Uppsala Genome Center for help with sequencing. We thank T. Bergström and L. Cavelier (Uppsala University) for providing RNA from chimpanzee spleen. We also thank A.I. Scollo, A.M. Perichon and M.C.R. Tenaglia (CEDIM, Diagnóstico Molecular y Forense SRL, Rosario, Argentina) for their help in DNA preparation of the Argentine samples and A. Voss for clinical help with the Danish samples. The authors would like to thank particularly the Lupus Patient Association of Asturias for their help in the collection of samples as well as all the patients for their contribution. The authors would also like to thank J.Osorio y Fortea (Pasteur Institut, Paris) for his contribution in the analyses that led to the identification of R61H early in this project.

This work has been supported in part by grants from the European CVDIMMUNE project from the European Commission LSHM-CT-2006-037227, Swedish Research Council for Medicine (12703), the Swedish Association against Rheumatism, the Magnus Bergwalls Foundation, the Gustaf V: 80th-year Jubilee Foundation, the Torsten and Ragnar Söderbergs Foundation and the Marcus Borgströms Foundation to M.E.A.-R. and by the Gurli and Edward Brunnbergs Foundation for rheumatologic research to S.V.K. M.E.A.-R. is supported by an award from the Knut and Alice Wallenberg Foundation through the Royal Swedish Academy of Sciences. This work was also partially supported by FISM, Regione Piemonte (CIPE), the Bundesministerium für Bildung und Forschung (BMBF) Kompetenznetz Rheuma C2.12, Germany and the Danish Rheumatism Association.

AUTHOR CONTRIBUTIONS
M.E.A.-R., S.V.K. and H.A. designed the experiments. S.V.K., A.-K.A., J.W., A.Z., M.V.P.L.R. and E.S. performed the experiments. M.E.A.-R., S.V.K., A.-K.A., J.W. and H.A. performed the analyses. I.G., E.S., G.S., L.T., A.J., T.W., S.D., N.B., M.G.D., C.G., A.S., P.J., H.L., B.A.P.-E., M.F.G.-E. and J.M. and their multicenter collaborators provided the samples. M.E.A.-R., S.V.K. and A.-K.A. wrote the manuscript.

The Argentine Collaborative Group participants are:
Hugo R. Scherbarth, Pilar C. Marino and Estela L. Motta (Servicio de Reumatología, Hospital Interzonal General de Agudos 'Dr. Oscar Alende', Mar del Plata, Argentina), Susana Gamron, Cristina Drenkard and Emilia Menso (Servicio de Reumatología de la UHMI 1, Hospital Nacional de Clínicas, Universidad Nacional de Córdoba, Córdoba, Argentina), Alberto Allievi and Guillermo A. Tate (Organización Médica de Investigación, Buenos Aires, Argentina), Jose L. Presas (Hospital General de Agudos Dr. Juán A. Fernandez, Buenos Aires, Argentina),

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

Simon A. Palatnik, Marcelo Abdala and Mariela Bearzotti (Facultad de Ciencias Medicas, Universidad Nacional de Rosario y Hospital Provincial del Centenario, Rosario, Argentina), Alejandro Alvarellos, Francisco Caeiro and Ana Bertoli (Servicio de Reumatología, Hospital Privado, Centro Medico de Córdoba, Córdoba, Argentina), Sergio Paira and Susana Roverano (Hospital José M. Cullen, Santa Fe, Argentina), Cesar E. Graf and Estela Bertero (Hospital San Martín, Paraná, Argentina), Cesar Caprarulo and Griselda Buchanan (Hospital Felipe Heras, Concordia, Entre Ríos, Argentina), Carolina Guillerón, Sebastian Grimaudo and Jorge Manni (Departamento de Inmunología, Instituto de Investigaciones Médicas 'Alfredo Lanari', Buenos Aires, Argentina), Luis J. Catoggio, Enrique R. Soriano and Carlos D. Santos (Sección Reumatología, Servicio de Clínica Medica, Hospital Italiano de Buenos Aires y Fundación Dr. Pedro M. Catoggio para el Progreso de la Reumatología, Buenos Aires, Argentina), Cristina Prigione, Fernando A. Ramos and Sandra M. Navarro (Servicio de Reumatología, Hospital Provincial de Rosario, Rosario, Argentina), Guillermo A. Berbotto, Marisa Jorfen and Elisa J. Romero (Servicio de Reumatología Hospital Escuela Eva Perón, Granadero Baigorria, Rosario, Argentina), Mercedes A. Garcia, Juan C. Marcos and Ana I. Marcos (Servicio de Reumatología, Hospital Interzonal General de Agudos General San Martín, La Plata, Buenos Aires, Argentina), Carlos E. Perandones and Alicia Eimon (Centro de Educación Médica e Investigaciones Clínicas (CEMIC), Buenos Aires, Argentina), and Cristina G. Battagliotti (Hospital de Niños Dr. Orlando Alassia, Santa Fe, Argentina). B.A.P.-E. is coordinator of the Argentine Collaborative Group.
The German Collaborative Group participants are:
K. Armadi-Simab and Wolfgang L. Gross (Abteilung Rheumatologie, University Hospital of Schleswig-Holstein, Campus Luebeck, Rheumaklinik Bad Bramstedt, Luebeck, Germany), Erika Gromnica-Ihle (Rheumaklinik Berlin-Buch, Berlin, Germany), Hans-Hartmut Peter (Medizinische Universitaetsklinik, Abteilung Rheumatologie und Klinische Immunologie, Freiburg, Germany), Karin Manger (Medizinische Klinik III derFAU Erlangen-Nuernberg, Erlangen, Germany), Sebastian Schnarr and Henning Zeidler (Abteilung Rheumatologie, Medizinische Hochschule Hannover, Hannover, Germany) and Reinhold E. Schmidt (Abteilung Klinische Immunologie, Medizinische Hochschule Hannover, Hannover, Germany).
The Spanish Collaborative Group participants are:
Norberto Ortego (Servicio Medicina Interna, Hospital Clínico San Cecilio, Granada, Spain), Enrique de Ramón (Servicio de Medicina Interna, Hospital Carlos Haya, Malaga, Spain), Juan Jiménez-Alonso (Servicio de Medicina Interna, Hospital Virgen de las Nieves, Granada, Spain), Julio Sánchez-Román (Servico de Medicina Interna, Hospital Virgen del Rocio, Sevilla, Spain) and Miguel Angel López-Nevot (Servicio de Inmunología, Hospital Virgen de las Nieves, Granada, Spain).
The Italian Collaborative participants are:
Gian Domenico Sebastiani (U.O.C. di Reumatologia Ospedale San Camillo, Roma, Italy), Enrica Bozzolo (IRCCS San Raffaele Hospital, Milan, Italy), Mauro Galeazzi (Siena University, Siena, Italy), Sergio Migliaresi (Rheumatology Unit Second University of Naples, Naples, Italy). Also we would like to thank Armando Gabrielli, Clinica Medica di Scienze Mediche e Chirurgiche, Universitá Politecnica delle Marche.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Sherer, Y., Gorstein, A., Fritzler, M.J. & Shoenfeld, Y. Autoantibody explosion in systemic lupus erythematosus: more than 100 different antibodies found in SLE patients. *Semin. Arthritis Rheum.* **34**, 501–537 (2004).
2. Sestak, A.L., Nath, S.K., Sawalha, A.H. & Harley, J.B. Current status of lupus genetics. *Arthritis Res. Ther.* **9**, 210 (2007).
3. Alarcon-Riquelme, M.E. The genetics of systemic lupus erythematosus: understanding how SNPs confer disease susceptibility. *Springer Semin. Immunopathol.* **28**, 109–117 (2006).
4. Forner, K., Guedj, M., Dauvillier, J. & Wojcik, J. Universal false discovery rate estimation methodology for genome-wide association studies. *Hum. Hered.* **65**, 183–194 (2008).
5. Yokoyama, K. *et al.* BANK regulates BCR-induced calcium mobilization by promoting tyrosine phosphorylation of IP(3) receptor. *EMBO J.* **21**, 83–92 (2002).
6. Aiba, Y. *et al.* BANK negatively regulates Akt activation and subsequent B cell responses. *Immunity* **24**, 259–268 (2006).
7. Burge, C.B., Tuschl, T. & Sharp, P.A. The RNA World II. (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 525–560 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
8. Purcell, S., Daly, M.J. & Sham, P.C. WHAP: haplotype-based association analysis. *Bioinformatics* **23**, 255–256 (2007).
9. Jordan, M.S., Singer, A.L. & Koretzky, G.A. Adaptors as central mediators of signal transduction in immune cells. *Nat. Immunol.* **4**, 110–116 (2003).
10. Kurosaki, T. Regulation of B-cell signal transduction by adaptor proteins. *Nat. Rev. Immunol.* **2**, 354–363 (2002).
11. Okada, T., Maeda, A., Iwamatsu, A., Gotoh, K. & Kurosaki, T. BCAP: the tyrosine kinase substrate that connects B cell receptor to phosphoinositide 3-kinase activation. *Immunity* **13**, 817–827 (2000).
12. Patterson, R.L., Boehning, D. & Snyder, S.H. Inositol 1,4,5-trisphosphate receptors as signal integrators. *Annu. Rev. Biochem.* **73**, 437–465 (2004).
13. Li, M. & Pritchard, P.H. Characterization of the effects of mutations in the putative branchpoint sequence of intron 4 on the splicing within the human lecithin:cholesterol acyltransferase gene. *J. Biol. Chem.* **275**, 18079–18084 (2000).
14. Mohler, P.J. *et al.* Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death. *Nature* **421**, 634–639 (2003).
15. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362 (1999).
16. Anolik, J., Sanz, I. & Looney, R.J. B cell depletion therapy in systemic lupus erythematosus. *Curr. Rheumatol. Rep.* **5**, 350–356 (2003).
17. Liossis, S.N., Kovacs, B., Dennis, G., Kammer, G.M. & Tsokos, G.C. B cells from patients with systemic lupus erythematosus display abnormal antigen receptor-mediated early signal transduction events. *J. Clin. Invest.* **98**, 2549–2557 (1996).
18. Huck, S., Le Corre, R., Youinou, P. & Zouali, M. Expression of B cell receptor-associated signaling molecules in human lupus. *Autoimmunity* **33**, 213–224 (2001).
19. Liossis, S.N. *et al.* B-cell kinase lyn deficiency in patients with systemic lupus erythematosus. *J. Investig. Med.* **49**, 157–165 (2001).
20. Hibbs, M.L. *et al.* Sustained activation of Lyn tyrosine kinase *in vivo* leads to autoimmunity. *J. Exp. Med.* **196**, 1593–1604 (2002).
21. Flores-Borja, F., Kabouridis, P.S., Jury, E.C., Isenberg, D.A. & Mageed, R.A. Decreased Lyn expression and translocation to lipid raft signaling domains in B lymphocytes from patients with systemic lupus erythematosus. *Arthritis Rheum.* **52**, 3955–3965 (2005).
22. Cornall, R.J. *et al.* Polygenic autoimmune traits: Lyn, CD22, and SHP-1 are limiting elements of a biochemical pathway regulating BCR signaling and selection. *Immunity* **8**, 497–508 (1998).
23. Kozyrev, S.V. *et al.* Structural insertion/deletion variation in *IRF5* is associated with a risk haplotype and defines the precise isoforms expressed in SLE. *Arthritis Rheum.* **56**, 1234–1241 (2007).
24. Tan, E.M. *et al.* The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **25**, 1271–1277 (1982).
25. Guedj, M., Wojcik, J., Della-Chiesa, E., Nuel, G. & Forner, K. A fast, unbiased and exact allelic test for case-control association studies. *Hum. Hered.* **61**, 210–221 (2006).
26. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162–1169 (2003).
27. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
28. Dudbridge, F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet. Epidemiol.* **25**, 115–121 (2003).

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

Corrigendum: Functional variants in the B-cell gene *BANK1* are associated with systemic lupus erythematosus

Sergey V Kozyrev, Anna-Karin Abelson, Jerome Wojcik, Ammar Zaghlool, MV Prasad Linga Reddy, Elena Sanchez, Iva Gunnarsson, Elisabet Svenungsson, Gunnar Sturfelt, Andreas Jönsen, Lennart Truedsson, Bernardo A Pons-Estel, Torsten Witte, Sandra D'Alfonso, Nadia Barrizzone, Maria Giovanna Danieli, Carmen Gutierrez, Ana Suarez, Peter Junker, Helle Laustrup, Maria Francisca González-Escribano, Javier Martin, Hadi Abderrahim & Marta E Alarcón-Riquelme

***Nature Genetics* 40, 211–216 (2008); published online 20 January 2008; corrected after print 27 March 2008**

In the version of this article initially published, the name of the 15th author was misspelled. The correct author name is Nadia Barizzone. Also, the affiliation of Javier Martin was incomplete. Dr. Martin is affiliated with Instituto de Biomedicina López-Neyra, Grenada 18100, Spain and Consejo Superior de Investigaciones Científicas (CSIC), Grenada 18100, Spain. The errors have been corrected in the HTML and PDF versions of the article.

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics