Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus

Simon H. Jiang1,2,3,21, Vicki Athanasopoulos1,2,21, Julia I. Ellyard1,2, Aaron Chuah3,2,4, Jean Cappello1,2, Amelia Cook1,2, Savit B. Prabhu1,2,5, Jacob Cardenas6, Jinghua Gu6, Maurice Stanley1,2, Jonathan A. Roco1,2, Ilenia Papa1, Mehmet Yabas1,7, Giles D. Walters2,1,2,3, Gaetan Burgio1,1, Kathryn McKeon1,2, James M. Byers1,2, Charlotte Burrin1, Anselm Enders1,2, Lisa A. Miosge1, Pablo F. Canete1,2, Marija Jelusic8, Velibor Tasic9, Adrian C. Lungu10, Stephen I. Alexander2,11, Arthur R. Kitching2,12, David A. Fulcher1,2,13, Nan Shen14, Todor Arsov1,2,14, Paul A. Gatenby13, Jeff J. Babon15, Dominic F. Mallon16, Carmen de Lucas Collantes17, Eric A. Stone18, Philip Wu2,19, Matthew A. Field2,4, Thomas D. Andrews2,4,20, Eun Cho2,4, Virginia Pascual6, Matthew C. Cook1,2,13,14 & Carola G. Vinuesa1,2,14

Systemic lupus erythematosus (SLE) is the prototypic systemic autoimmune disease. It is thought that many common variant gene loci of weak effect act additively to predispose to common autoimmune diseases, while the contribution of rare variants remains unclear. Here we describe that rare coding variants in lupus-risk genes are present in most SLE patients and healthy controls. We demonstrate the functional consequences of rare and low frequency missense variants in the interacting proteins BLK and BANK1, which are present alone, or in combination, in a substantial proportion of lupus patients. The rare variants found in patients, but not those found exclusively in controls, impair suppression of IRF5 and type-I IFN in human B cell lines and increase pathogenic lymphocytes in lupus-prone mice. Thus, rare gene variants are common in SLE and likely contribute to genetic risk.
Systemic lupus erythematosus (SLE) is a highly heterogeneous autoimmune disease in terms of both clinical phenotypes and underlying pathophysiology. Examination of disordered immune function in patients and animal models of lupus have implicated a diverse range of mechanisms contributing to disease. B-cell hyperreactivity and breaks in B-cell tolerance caused by B-cell intrinsic and extrinsic factors often involving TLR7 signaling are central to disease pathogenesis. The most consistent final common feature in human SLE patients is overproduction of type I interferon (T1 IFN). Monogenic interlinkage disequilibrium with GWAS-identified variants as a plausible explanation for pathogenicity. The prevalent hypothesis is that the cumulative or epistatic effect of multiple GWAS single-nucleotide polymorphisms (SNPs) in addition to environmental factors that result in predisposition to SLE. Lupus is often a devastating disease in women of childbearing age, which might argue against the aggregate contribution of common genetic variants. Another hypothesis is that in some cases SLE arises from rare genetic variants with strong effects. Rare (MAF < 0.005) and low frequency (MAF < 0.02) variants are known to contribute to complex hereditary traits and can explain sporadic disease in the case of de novo mutations. These disease-predisposing rare variants can be in linkage disequilibrium with GWAS-identified SNPs due to co-inheritance in shared haplotypes and may contribute to the missing heritability in common autoimmune diseases.

In this study we examine the role of rare and low-frequency variants in systemic autoimmunity. We identified several rare variants in BLK and BANK1 in healthy controls and patients with SLE. BLK variants found in SLE patients impaired the kinase activity of BLK. We demonstrate that BLK is capable of repressing IRF5-mediated interferon-β expression (IFNβ), and that loss of BLK kinase activity caused by the rare variants enhanced interferon-β production. In contrast rare variants in BLK found exclusively in healthy controls did not substantially impair interferon-β repression. As expected, SLE patients with rare BLK variants also had increased expression of interferon signature genes compared to healthy controls. We observed that low-frequency mutations in BLK’s epistatic partner BANK1, impair BANK1’s ability to sequester IRF5 from the nucleus. Together these data identify a novel function for BANK1 and BLK in repression of type I IFN and demonstrate how rare variants in SLE risk genes increase type I interferon activity which is central to development of autoimmunity.

**Results**

SLE patients have rare coding variants in SLE-risk genes. To investigate the prevalence of rare variants in lupus-risk genes we selected an initial 69 SLE probands (SLE1), a replication cohort of 64 SLE probands (SLE2) and 97 healthy elderly individuals without a history of chronic disease. The SLE cohort comprised pediatric-onset (14%), adolescent-onset (26%), and adult-onset (55%) SLE and four patients whose age of disease onset was unknown. The genetically determined ethnicity of the two cohorts was predominantly European with 79.7% of SLE patients and 100% of healthy controls of European ethnicity.

We generated a list of all identified lupus-associate genes including all monogenic causes of SLE, known causes of interferonopathies and many of the variants in systemic lupus erythematosus (SLE) genome association studies (GWAS). We identified several rare variants in SLE-risk genes as did 72% of SLE1 and 63% of SLE2 patients. Of the 76 genes, there were 20 genes with variants in our cohort. Of the 76 genes, there were 20 genes with variants in our cohort. Of the 76 genes, there were 20 genes with variants in our cohort. Of the 76 genes, there were 20 genes with variants in our cohort. Of the 76 genes, there were 20 genes with variants in our cohort. Of the 76 genes, there were 20 genes with variants in our cohort.

**Variants in BLK and BANK1**

We next asked whether the rare variants observed in patients with SLE confer functional defects. For this we asked which lupus-associated genes harbour the rare variants in our cohort. Of the 76 genes, there were 20 genes with variants in ≥3 SLE patients (2.3%) and a prevalence ≥1.5 times that seen in controls (Fig. 1c, Table S2). BLK, LYST, TYK2, UHRF1BP1, and IKBKE were the genes most frequently containing rare variants in SLE patients. To test the contribution of rare variants to disease, we next turned to functional studies.

We started by looking at the SNV in BLK given the high frequency of rare variants in this gene. BLK was also chosen because multiple common variants tagging BLK (rs2248932, MAF: 0.49; rs2736340, MAF: 0.38; and rs13277113, MAF: 0.36) have been identified as SLE-predisposing by GWAS (OR = 1.39 [95% CI: 1.28–1.51]). Also, BLK in humans is expressed in B cells and plasmacytoid dendritic cells, two cell types considered to be important for SLE development. The BLK GWAS variants reported to date are not coding and suggested to overlap with regulatory elements leading to reduced BLK gene expression. In the initial SLE cohort we identified 6 rare or novel SNV in BLK (BLKR1, rs73663163, BLKR13Q, rs144615291, BLKR25Q, rs141865425, BLKp507T, novel, BLKp507H, rs758750492, and BLKR598, rs146505280, and BLKp598, rs146505280) in 14 patients and an additional low-frequency variant BLK171T (rs55758736, MAF: 0.012) in 5 patients (Fig. 2a and Table 1). Thus, 10.1% of our original SLE patients (SLE1, n = 7 of 69) had a rare or novel missense SNV in BLK. Rare or novel missense SNVs in BLK were found at significantly lower frequencies in a...
common variable immunodeficiency/complex immunodeficiency (CVID/CID) cohort \((n = 3 \text{ of } 107, 2.8\%, p \leq 0.02)\) but at comparable frequency in healthy controls \((n = 7 \text{ of } 97, 7.2\%, p = 0.5)\). Synonymous SNV of any allelic frequency in BLK were found at equivalent rates in all three cohorts (Fig. 2b).

All relatives with systemic autoimmunity in the pedigrees of probands with BLK rare alleles also carried the BLK variants, however, 25% of individuals carrying the variants did not have disease. While there are multiple factors shown to contribute to incomplete gene penetrance (gender, environmental influences, epigenetic changes, combined effects of risk and protective common alleles, etc.), we considered the possibility of additional SLE rare variants acting in epistasis with BLK being present in subjects with disease. In search for variants in BLK-interacting proteins, we identified 2 families in which there was cosegregation in patients with autoimmune disease of a rare BLK SNV with a rare and a low-frequency mutation in BANK1 (Supplementary Fig. 2); both SNVs occur within exon 2.

We hypothesised that, like the BANK1R61H/G variant, impaired kinase activity of BLK by impairing availability of protein or protein stability and thus minimize available protein 32. Indeed, when we searched for this variant in a third cohort comprised of 150 SLE patients (for which the DNA yield and quality was insufficient for WES, but adequate for a targeted polymerase chain reaction (PCR)-amplifier assay), we identified the BANK1W40C SNV in 9 SLE probands (6%) and in 3 out of 222 unaffected controls \((n = 3 \text{ of } 222, 1.3\%)\) resulting in an OR of 4.7 \((95\% \text{ CI: } 1.1–27.1, p = 0.017)\) (Fig. 2d). No further low-frequency or novel BANK1 variants were found in the SLE cohort (Fig. 2e).

Together these data demonstrate that a substantial fraction of SLE patients have one of multiple rare- or low-frequency SNVs in BANK1 and/or BLK.

**BLK SNVs impair phosphorylation of BANK1 and IRF5.** We next examined the effect of the BLK and BANK1 SNVs on protein function. We noted that many of the BLK SNVs identified are in functionally important regions of the protein. The arginine residue at position 238, mutated in families D and G, is strictly conserved in all Src-family kinases. In the resolved protein crystal structures this residue is seen to orient the SH2-kinase linker with the N-lobe of the kinase domain via an interaction with P30731, which is mutated in family N (Fig. 3a). Mutation of the analogous R238 residue in Src has been shown to inhibit its catalytic activity31. In addition, the arginine residue at position 131, mutated in families A and C, is conserved in many SH2 domains and assists in the coordination of bound phosphotyrosines. The tyrosine at 350, mutated in family J, resides within the kinase domain and we hypothesized would restrict BLK catalytic activity. BLK771T is known to reduce protein stability and thus minimize available protein32. Therefore, we postulated that the identified SNVs would alter kinase activity of BLK by impairing availability of protein or protein function. BLK expressed in heterologous cell lines translates a higher molecular weight active band due to relatively higher phosphorylation, and a lower molecular weight inactive band38. As expected, upon transfection of BLKR131Q, BLK771T, BLK133W, and BLK238Q into HEK293T cells almost no active bands were seen, compared with visible active bands in cells transfected with wild type BLK, BLK133W, and a constitutively active BLK771F (Fig. 3b). This suggests that some of the BLK mutations impair the protein’s ability to acquire an active

![Fig. 1 Large burden of rare nonsynonymous SNV in SLE-related genes in SLE patients. a Pie chart demonstrating burden of variants in two SLE cohorts, SLE1 and SLE2, and 97 elderly healthy controls (HC). The numbering 0–4 indicates the number of rare variants found in a list of 76 GWAS SLE genes. b Number of SLE GWAS risk alleles in individual SLE patients with indicated number of rare variants (mean and SD, Kruskal–Wallis test). c The most frequently mutated genes in SLE cohort of 113 probands, shown as the number of total SLE patients and healthy controls (proportion of subjects) with variants in each individual gene.](image)
conformation. As BLK is known to phosphorylate BANK1 and SNVs were identified in both genes in the three patients, we tested whether the identified BLK variants had impaired ability to phosphorylate BANK1. Indeed, the four rare BLK variants tested had impaired phosphorylation of BANK1 (Fig. 3c). Collectively, this demonstrates that rare variants in BLK identified in a large proportion of SLE patients have significantly impaired kinase function.

**BLK SNV impair repression of IRF5-mediated T1 IFN expression.** We then explored the contribution of rare alleles of BLK to the pathogenesis of SLE. Since BLK activates PLCG2 and induces Ca\(^{2+}\) flux upon B-cell receptor (BCR) signaling\(^34\), we tested whether the variant may impair Ca\(^{2+}\) flux in a human B-cell line (Ramos) in which we introduced the BLK\(^{R131W}\) variant using CRISPR-Cas9 editing. No difference in response to BCR-stimulation of cells compared with WT cells, excluding this as a potential mechanism (Supplementary Fig. 3).

BLK is a member of the SrcB kinase subfamily, which includes LYN. LYN and BLK are both activated by CD79-mediated phosphorylation upon BCR cross-linking\(^35\), and both...
Immunity through impaired regulation of IRF5-mediated T1 IFN

LYN

of CRISPR/Cas9-engineered dependent manner (Supplementary Fig. 4). In addition, using a IFNb activity compared to wild-type BLK (Fig. 4a) in a dose-BLKR238Q

de
Lyn

regulating calcium

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-10242-9 ARTICLE

phosphorylate BANK1 after BCR stimulation34. Common alleles of LYN, like BLK, have been implicated in SLE by GWAS13 and Lyn deficiency induces a lupus-like phenotype in mice36. Besides regulating calcium flux36, Lyn deficiency also promotes autoimmunity through impaired regulation of IRF5-mediated T1 IFN production and activity11. We thus hypothesized that BLK may share IRF5 as a substrate with LYN and the hypomorphic BLK variants may impair IRF5 regulation. Indeed, upon co-expression of BLK and IRF5 in HEK293T cells, wild-type BLK phosphorylated IRF5, whereas the identified rare BLK variants had diminished or no ability to phosphorylate IRF5 (Fig. 3d).

We next tested whether, as shown for LYN, BLK plays an active role in the repression of IRF5 and T1 IFN activity. Expression of the BLK variants with an IFNb dual luciferase reporter demonstrated that all tested BLK variants (BLK R131W, BLKR238Q, and BLK V350F) were unable to repress IRF5-mediated IFNb activity compared to wild-type BLK (Fig. 4a) in a dose-dependent manner (Supplementary Fig. 4). In addition, using a CRISPR/Cas9-engineered BLK R131W/IR31W human B-cell line we demonstrated enhanced IFNb expression in response to stimulation with the TLR7/8 agonist resiquimod (R848) (Fig. 4b).

We also asked whether the rare BLK variants found in SLE patients were more damaging than those found in healthy controls. For this we generated plasmids expressing each of the six rare BLK variants exclusively found in healthy controls (BLK R131W, BLK H55R, BLK R84N, BLK V101H, BLK R115Q, BLK 5248A Table 2) and tested side-by-side the repressive ability of all BLK variants. Five of six rare BLK variants found in patients with SLE had greater than 50% reduction in IFNb repression (Fig. 4c) compared to wild-type BLK, whereas none of the six rare BLK variants found exclusively in healthy controls had a similar impairment (p < 0.0001). Consistent with the inability of SLE patient BLK variants to repress IFNb, examination of peripheral blood mononuclear cells from family G patients heterozygous for the BLK R238Q variant (G.1,1, G.1,2, and G.1,3) revealed upregulation of the T1 IFN signature (Fig. 4c and module M1.2 in Fig. 4d) as well as apoptosis/survival pathways (module M6.6 in Fig. 4d, Supplementary Fig. 5). Together these data demonstrate that rare BLK variants in patients with SLE have impaired regulation of IFNb expression whereas those found only in healthy controls do not.

Variants in BLK augment pathogenic T cells in Faslpr mice. We next tested the effect of the BLK variants on lupus development in vivo using CRISPR/Cas9-generated mice bearing the orthologue of human BLK R131W, Blk R125W (Fig. 5a). Blk R125W/R125W mice had normal immune phenotypes comparable to those of heterozygous and wild type littermates (Supplementary Fig. 6A). Unlike the human BLK R131W variant, stimulation of lymphocytes from CRISPR/Cas9-engineered mice expressing Blk R125W did not reveal an increased T1 IFN response (Supplementary Fig. 6B) consistent with previous studies suggesting redundancy between Blk and other Srf kinases in mice37. The Blk R125W variant was introduced into mice with a genetic susceptibility to SLE. We chose C57BL/6, Faslpr mice because Faslpr mice in the MRL genetic background develop a syndrome that resembles human lupus and Blk haploinsufficiency exacerbates lupus in MRL Faslpr mice26. B6.Faslpr mice carrying a single Blk R125W allele showed significantly expanded CD4/CD8 double-negative T cells (Fig. 5b), which have been shown to be important contributors to the hypercellularity in these

Table 1 Comparison of BANK1 and BLK variants found in SLE patients

| Variant | MAFa | Type | AA change | CADD | HC with allele (n) | SLE with allele (n) |
|---------|------|------|-----------|------|-------------------|-------------------|
| rs2248932 | 0.61 | Intron | N/A | N/A | – | – |
| rs13277113 | 0.25 | Intron | N/A | N/A | – | – |
| rs2736340 | 0.25 | Intron | N/A | N/A | – | – |
| rs55758736 | 0.01 | Missense | A7T | 5.86 | 1 | 7 |
| rs73663163 | 0.0005 | Missense | R31W | 14.93 | 0 | 1 |
| rs144615291 | 0.0004 | Missense | R31Q | 16.05 | 0 | 1 (n = 1 hom) |
| rs141865425 | 0.003 | Missense | R238Q | 22.2 | 2 | 3 |
| N/A | Novel | Missense | P307R | 21.8 | 0 | 1 |
| rs177401687 | 0.002 | Missense | K235T | 18.18 | 0 | 1 |
| rs758750492 | 0.0000004 | Missense | Y350H | 15.7 | 0 | 2 |
| rs146505258 | 0.0008 | Missense | R359C | 20.1 | 1 | 1 |
| N/A | Novel | Missense | R433Q | 17.35 | 0 | 1 |
| rs20162624 | 0.002 | Missense | R450H | 18.23 | 0 | 2 |

*Minor allele frequency by gnomAD

GWAS genome wide association study, NRLF novel, rare and low-frequency, Hom homozogous, HC healthy control, SLE systemic lupus erythematosus

Table 2 Rare nonsynonymous BLK variants found in healthy controls

| Variant | MAFa | Type | AA change | CADD |
|---------|------|------|-----------|------|
| rs142352008 | 0.002 | Missense | P39L | 22.7 |
| rs202053568 | 0.0003 | Missense | H55R | 5.8 |
| rs778435147 | 0.0006 | Missense | L84N | 26.7 |
| rs371256341 | 0.0002 | Missense | V101I | 5.2 |
| rs367628135 | 0.0002 | Missense | R115Q | 25 |
| rs763307492 | 0.000004 | Missense | G248A | 24.1 |

*Minor allele frequency by gnomAD
mice and to provide help for autoantibody production in humans. Together, these data suggest that human BLK phosphorylates IRF5 and the hypomorphic BLK alleles have impaired phosphorylation of IRF5 and diminished IRF5-mediated T1 IFN repression. Furthermore, a BLK allele orthologous to the one found in SLE was found to contribute to disease in lupus-prone mice.

**BANK1 SNV enhances TRAF6-mediated nuclear IRF5 localization.** BANK1, initially identified as a substrate of LYN, is a scaffold protein lacking intrinsic kinase activity. Scaffolds may positively and negatively regulate intracellular signalling pathways in innate and adaptive responses by controlling availability and post-translational modification of signalling proteins. When expressed in HEK293T cells BANK1 formed cytoplasmic inclusion bodies in a proportion of cells reminiscent of TRAF6-containing sequestosomes. Indeed, we confirmed wild type BANK1 colocalized with TRAF6, the sequestosome protein p62, as well as with the deubiquitinating enzyme CYLD, which plays a critical role in regulating TLR signaling by deubiquitinating TRAF6 (Supplementary Fig. 7A). BANK1 also localized in cytoplasmic aggregates when co-expressed with MYD88 (Supplementary Fig. 7B). We confirmed that BANK1 formed a complex with TRAF6 by coimmunoprecipitation upon expression in HEK293T cells (Fig. 6c).

TRAF6 is known to ubiquitinate IRF5 resulting in activation and subsequent nuclear localization of IRF5 and T1 IFN production. We thus hypothesized that BANK1 promotes sequestration of TRAF6 in typical p62+ and CYLD+ sequestosomes, reducing TRAF6 ubiquitination and thereby dampening IRF5 activation and induction of T1 IFN (Fig. 7a). Consistent with this hypothesis, expression of BANK1W40C leads to significantly reduced formation of BANK1+ sequestosomes (Fig. 7b, c). Furthermore, when expressed with TRAF6 and IRF5, WT BANK1 significantly repressed TRAF6-mediated IRF5 nuclear localization. By contrast, BANK1W40C could not repress TRAF6-mediated IRF5 nuclear localization to the extent of WT BANK1 (Fig. 7d). These findings demonstrate that the scaffold protein BANK1 regulates TRAF6 activity and hence IRF5 signaling, and establish that BANK1W40C is a loss-of-function variant that promotes T1 IFN activity.
Fig. 4 BLK variants fail to repress type 1 IFN expression. a IFNβ luciferase activity 24 h after co-transfection of HEK293T cells with IRF5, MyD88, and indicated BLK variants. Each data point represents the relative luciferometer unit (RLU) relative to a cotransfected control plasmid expressing renilla luciferase, and is the average of three technical replicates. (*p < 0.05, **p < 0.01, ***p < 0.001 Student t test). b IFNβ luciferase expression in CRISPR-Cas9 edited Ramos cells homozygous for the R131W variant, 24 h after stimulation with resiquimod (R848). Each data point represents the relative luciferometer unit (RLU) relative to a cotransfected control plasmid expressing renilla luciferase, and is the average of five technical replicates. (*p < 0.05, **p < 0.01, Mann-Whitney U). c IFNβ luciferase activity relative to WT BLK 24 h after co-transfection of HEK293T cells with IRF5, MyD88, and indicated BLK variants found in healthy controls (HC) or SLE patients (SLE). Data points for each variant were normalized to the repression levels of WT BLK. The dotted line indicates 50% of repression by WT BLK (mean and SD). d Heat map of 34 hierarchically clustered genes within IFN module M1.2. The data are normalized to healthy controls by subtracting out the mean of the log ratios of the genes. The map shows the percentage of genes within each module (minimum 10%) that are significantly upregulated and 25% are significantly downregulated would appear on the map as 50–25% = 25% upregulated and would have a light red color.

Discussion

Although SLE is the result of a combination of environmental and intrinsic predispositions, genetic risk remains one of the most potent risk factors. GWAS have provided substantial advances in identifying possible disease pathways, yet they have been less informative about disease mechanisms. This is because variants identified by GWAS, which are typically found at high frequencies, only modestly increase risk and in the significant majority of cases have modest or no effect on protein function. The prevalent hypothesis to explain substantial genetic risk for SLE with common, weak variants is that risk arises from the cumulative burden of dozens of these GWAS alleles. However, the expanding, but still small, list of monogenic causes of SLE supports the notion that novel or rare gene variants that significantly cripple the function of crucial DNA-sensing or degrading enzymes, or complement factors involved in the...
removal of apoptotic cells, can cause severe and often early onset SLE-like disease\textsuperscript{20}. We show here that SLE patients are likely to harbour two or more rare variants in genes implicated in SLE by GWAS or involved in the regulation of T1 IFN. The main focus of the study was to determine whether the variants found in SLE patients were damaging and could contribute to disease, and if so, whether they were more damaging than those found in controls.

We tested the functional consequences of mutations in two genes: BLK, which harboured novel or rare variants in 10.5\% of SLE probands, and BANK1, encoding a known BLK interacting partner and shown to act in epistasis with BLK\textsuperscript{30}. Significantly, we demonstrate that these rare variants exert measurable damaging effects on protein function, ultimately leading to a common endpoint of increased T1 IFN activity in human B cells. Excessive T1 IFN activity is a unifying feature in the majority of SLE patients\textsuperscript{44}. Moreover, since BANK1 and BLK are also expressed in plasmacytoid dendritic cells\textsuperscript{26,28,29}, it is likely that a similar failure to repress T1 IFN occurs in these cells, which are major producers of these cytokines\textsuperscript{45}. Thus, these data indicate that rare SNVs in the GWAS-implicated genes BANK1 and BLK are associated with development of lupus and related autoimmune diseases.

We identify a novel role for BLK in regulating T1 IFN downstream of TLR7/8 signalling, and demonstrate that the rare variants found in SLE patients impair the ability of BLK to repress T1 IFN production. The incomplete penetrance of autoimmune in BLK heterozygous individuals and the absence of overt autoimmunity in BLK\textsuperscript{R125W/R125W} mice may be due to the differences in environmental exposure to stimulants of T1 IFN such as viral infection. It is also likely that there is increased redundancy of Blk in mice, in which other Src family-kinases have been shown to compensate for Blk deficiency\textsuperscript{37}. Moreover, the presence of additional lupus-predisposing gene variants in SLE patients may enhance the BLK defects. Strikingly, we observed a distinct difference in the deleteriousness of rare variants in BLK in HC and SLE, suggesting that quality (degree of damage to protein function) rather than quantity (number of rare variants) may be a more important determinant of contribution of rare variants to disease.

BANK1, a scaffold protein, lacks intrinsic kinase activity and may regulate signalling events by localisation or sequestration of significant intracellular signalling proteins. Here, we show that a low-frequency (MAF < 2\%) BANK1 variant diminishes localisation of BANK1 to sequestosomes, likely altering recruitment of TRAF6 and CYLD to these regulatory structures. This is associated with enhanced nuclear localization of IRF5, which is a positive regulator of T1 IFN transcription\textsuperscript{46}. Since BANK1 is a direct target of BLK phosphorylation, and BANK and BLK have been previously shown to act in epistasis\textsuperscript{30}, BLK’s capacity to repress T1-IFN is probably related to its ability to regulate BANK1 localization to sequestosomes and as a consequence, regulate sequestosome homeostasis and TRAF6 activity.

Our demonstration that rare gene variants in GWAS-identified SLE risk genes are damaging, and that these occur in a large fraction of SLE patients supports the notion that rare variants, as shown in common and genetically complex conditions\textsuperscript{16,17}, contribute to SLE pathogenesis. Furthermore, the co-segregation of variants in BANK1 and BLK in two families suggests that sporadic lupus may occur upon inheritance or de novo occurrence of two or more rare variants with strong effects that act together to cause disease.

Together, our findings demonstrate a role of rare BLK and BANK1 variants in SLE and may offer an alternative explanation for the association of some common variants in linkage disequilibrium. Identification of rare variants as causes of lupus and related systemic autoimmune disorders through whole exome sequencing provides an approach intermediate between GWAS and conventional mapping. An approach that demands functional verification, but if provided, as we show here, can also yield novel insights into disease mechanisms and novel targets for treatment.

Methods

Human patients and DNA sequencing. Written informed consent was obtained as part of the Australian Point Mutation in Systemic Lupus Erythematosus study (APOSLE) and the Centre for Personalised Immunology program. The study was approved by and complies with all relevant ethical regulations of the Australian National University and ACT Health Human Ethics Committees. SLE cohort 1 and 2 were both recruited in Australia, processed and sequenced on similar platforms. These two cohorts were recruited sequentially, with SLE cohort 1 recruited between 2008 and 2014 and SLE cohort 2 recruited between 2015 and 2017. Saliva was collected in Oragene” DNA self-collection kits and purified using PrepIt” DNA purification kits (Oragene) and treated with Ribonuclease A (Qiagen Cat# 19101). DNA samples were enriched using Human SureSelect XT2 All Exon V4 Kit and sequenced by Illumina HiSeq 2000 (Illumina, Inc.). WES had 21\% low or uncov    ered exon pairs compared with 4\% low or uncovered exons for WGS. Bioinformatic analysis was performed at JCSMR, ANU. Raw sequence reads were aligned to the reference genome (Hg19) and single-nucleotide variants and small insertions and deletions called using GATK. Results were scored based on reported minor allelic frequency (MAF), Polyphen2 score, expression in immune tissues and reported mouse phenotypes. All SNVs of interest in BLK and BANK1 were confirmed by Sanger sequencing. Amplifiable to detect BANK1\textsuperscript{39,46} and BLK\textsuperscript{32,39} in the APOSLE cohort was performed using the CHEMICON Amplifiier SNPs TTT Genotyping System Fam-Joe kit S7909 (Merk-Millipore). The 45 and Up\textsuperscript{47,48} and ASPREE\textsuperscript{49} datasets were used as reference healthy controls, accessed through the MGRB Collaborative (http://sgc.garvan.org.au/mgrb/initiatives).

WES/WGS data processing and batch correction. Probes were filtered out if the detection p value was greater than 0.01 for at least 100\% of the samples. All data values <10 were set to 10 and then the data was log2 transformed. An additional filter selecting the 75\% most variable transcripts was performed, leaving a total 18,004 probes for analysis. Principal variance component analysis (PVCA) was
conducted to identify undesirable sources of technical variability within the data and batch correction was applied to correct for this technical variation. Both PVCA and batch correction were conducted using IMP Genomics 7.0 (SAS Institute) analysis software.

**Determination of ethnicity by WES/WGS.** We determined each individual’s ethnicity utilizing GEMTools with genotypes across 23,536 sites extracted from all 2504 Phase 3 samples of the 1000Genomes project to cluster our 230 samples individually into corresponding 1000Genomes superpopulations which directly correspond to 6 populations within gnomAD (AFR, AMR, EAS, FIN, NFE, and SAS). In all 230 GEMTools clustering runs (after 24 h with non-default parameter was setting maximum individual cluster size to 114 to match the maximal sample size of 1000Genomes’ 26 populations), each of our samples fell unambiguously within a population-homogenous cluster which was assigned as its best-matching population. Admixture was determined in our sample cohorts using ADMIXTURE, an implementation of the ADMIXTURE algorithm to correct for population stratification, applied on the Dodecad K7b (http://dodecad.blogspot.com/2012/01/k12b-and-k7b-calculators.html) global ancestry reference panel. This method, though accurate for determining ancestral population components across continents, was not suited for binning our samples into an admixed gnomAD population like AMR, thus we used our GEMTools-derived best-matching population to select each sample’s ethnically matched gnomAD population frequency.

**RNA-expression analysis.** Whole blood was collected in acid citrate dextrose (ACD) tubes. RNA was extracted from whole blood (5’ Prime Perfect Pure kit) and stored at ~80°C until use. Differential gene expression analysis was performed using linear modeling with the Limma package. Gene-set analysis was conducted using the QuSAGE algorithm, which tests whether the average log2-fold change of a gene set is different from 0 and takes into account the correlations of the genes by incorporating an estimate of the variance inflation factor of the gene set. Module maps were generated as reported previously.

**Expression vectors and mutagenesis.** The following expression vectors were obtained: untagged BLK (OriGene Technologies), myc-LYN (GeneCopeia, Inc.), untagged IRF5 (OriGene Technologies). BANK1-V5 was a gift from C. Castillojo-Lopez (Uppsala University). Mutagenesis was performed using the Quikchange II and II Site directed mutagenesis protocols (Agilent Technologies). IFNβ luciferase (Addgene) pRL-CMV (Promega). pCMV-HA-MyD88 was a gift from Bruce Beutler (Addgene plasmid # 12287), pX330-U6-Chimeric_BB-CBh-hSpCas9 was a gift from Feng Zhang (Addgene plasmid # 42230).

**Antibodies.** Antibodies for western blotting, immunofluorescence imaging, and communoprecipitation studies were as follows: mouse anti-human BLK (SC-65990; Santa Cruz), Mouse anti-V5 (clone SVS-PK1, MCA1380, BioRad); mouse anti-FLAG M2 (F1804; Sigma); Rabbit anti-LYN (06-207, EMD Millipore); mouse anti phospho tyrosine-horseradish peroxidase (HRP) (R&D Systems); IRF5 (Abcam) and monoclonal mouse anti-CD71 (transferrin receptor) antibody (C2063, Sigma). Secondary antibodies were conjugated to HRP (Jackson ImmunoResearch), Alexa 568 or Alexa 488 (Molecular Probes, Invitrogen). Indo-1 AM was from ThermoFisher and anti-IgM from Jackson ImmunoResearch. All FACS and microscopy work was carried out at the Microscopy and Cytometry Facility, Australian National University.

**Flow cytometry.** The study was approved by and mouse handling complies with all relevant ethical regulations of the Australian National University Ethics Committee. Spleens were isolated as single-cell suspensions after red blood cell lysis. To stain cell surface markers, we incubated cells in the antibody mixture diluted in ice-cold staining buffer (2% fetal calf serum in phosphate-buffered saline). Ramos cells were loaded with Indo-1 AM at 37°C for 2 h before being stimulated at 37°C with anti-Fab(2) antibody. An LSRII or Fortessa Flow Cytometer with FACSDiva software were used for flow cytometry acquisition, and FlowJo (Tree Star) was used for analysis.

**Transfection, communoprecipitation, and western blotting.** HEK 293T cells were transfected (Lipofectamine 2000, Life Technologies) with the relevant plasmids as per manufacturer’s recommendation. Cells were lyzed using NP-40 lysis buffer and immunoprecipitated with the relevant antibody using Protein G Sepharose (GE Healthcare) and the relevant antibody. For communoprecipitation experiments transferrin receptor was used as isotype control. Immunoprecipitants were resuspended in SDS (sodium dodecyl sulfate) buffer and boiled prior to electrophoresis on 8% SDS-polyacrylamide gel electrophoresis gels. Gels were transferred to nitrocellulose membranes (BioRad Laboratories), blocked overnight (TBST + skim milk powder; or 5% bovine serum albumin for phosphotyrosine blots) and probed with the relevant primary and secondary antibodies. Membranes were developed with chemiluminescence developer (Western Lightning Plus ECL; Perkin Elmer).

**DLAs and electroporation.** HEK293T cells were transfected with an IFN-β luciferase reporter, pRL-CMV (10 ng; Promega) Renilla luciferase control reporter, pCDNA 3.1 and indicated vectors and 24 h later dual luciferase assays (DLAs) were performed as per published protocols. For CpG response assays, TLR9-HEK293s (Invivogen) were transfected using lipofectamine with indicated vectors, IFNβ-luciferase and renilla reporters before being stimulated after 24 h with 5 μg/ml CpG (ODN 2006, Invivogen). The Burkitt’s lymphoma cell line Ramos was nucleasefed with the IFN-β luciferase reporter and pRL-CMV using the NEON
transfection system (ThermoFisher Scientific) according to the manufacturer’s instructions.

CRISPR-Cas9 genome editing of human B-cell lines. The vector px33062 (Addgene), which expresses Cas9 and the sgRNA, was linearized with BsrI and gel-purified. A pair of complementary 18mer oligos targeting a single site within the genome (primer sequences available upon request) were annealed and ligated to the linearized vector. px330 expressing the sgRNA of interest was transfected along with a mcherry plasmid into the Burkitt’s lymphoma cell line Ramos using the NEON transfection system (ThermoFisher Scientific) according to the manufacturer’s instructions. mcherry+ single cells were sorted 24–48 h later into 96-well plates and individual clones expanded until confluent. gDNA was extracted by digesting cells with proteinase K (0.1% Tween20, 100 µM EDTA, 500 µg/mL). Proteinase K, 1× high fidelity Phusion buffer) at 56 °C for 40 min, followed by incubation at 95 °C for 8 min. A region of approximately 300 bp flanking the variant was PCR amplified (primer sequences available on request) using Phusion DNA Polymerase II (ThermoFisher Scientific) and presence of the mutation confirmed by Sanger sequencing. All sequencing was carried out at the ACRF Biomolecular Resource Facility and Genome Discovery Unit, Australian National University.

CRISPR-Cas9-mediated genome editing of mouse zygotes. C57BL/6 mice were housed under specific pathogen-free conditions. All mouse procedures had been approved by the Australian National University Animal Experimentation Ethics Committee. (AEEC A2014/058 and A2014/016) under the NHMRC Australian code of practice. B6 gRNA and Cas9 protein were obtained from PNABio. Oligo and ssOligos were purchased from IDT (sequences available on request). C57Bl/6Ncrl female mice (4–5 weeks old) were superovulated with Pregnant Mare Serum Gonadotrophin (PMSG) SUI day 1 and Human Chorionic Gonadotrophin hormone (HCG) SUI day 3. After detection of a vaginal plug of the superovulated females, mouse zygotes were harvested from the ampulle and were placed in KSOM medium (Sigma). Cas9 protein (100 ng/µl) was co-injected with a mixture of sgRNA (50 ng/µl each) and ssOligo (100 ng/µl) into the cytoplasm of the fertilized eggs into M2 medium (Sigma). After micro-injection, the zygotes were incubated overnight at 37 °C and 5% CO2 and two-cell stage embryos were surgically transferred into the uterus of pseudopregnant CD1 recipient females at 2.5 dpc. Three weeks after birth mouse ears were punched. DNA was extracted and Sanger sequencing performed to confirm the mutations. All the mouse zygote preparation and micro-injection was carried out at the Australian Phenomics Facility, Australian National University. The sequencing was carried out at the ACRF Biomolecular Resource Facility and Genome Discovery Unit, Australian National University.

qPCR. Total RNA was extracted from splenocytes stimulated in the presence of R848 for 24 h using Trizol (Invitrogen). cDNA was synthesized using superscript III (Thermo) or miscript (Qiagen) according to manufacturer’s instructions. Quantitative PCR (qPCR) was carried out using the SYBR green method with the following primer sequences: Eif2ak1 (forward, 5′-ATGCAGGGATAGCATTTAGCA-3′) and (reverse, 5′-AGTTCCTACTGACCCA-3′); reverse-GACAATCCACCTTGTTTGTG-3′), If44 (forward, 5′-AACTGACTGTCGGAAATAATTG-3′); reverse-AGCAATCGCCCTTGGTGT-3′, Igs15 (forward, 5′-GGTGCCTGAATCTTGCAT-3′; reverse-TGGAAAGGTAAGACCGTCCT-3′), JIF7 (forward, 5′-GAACATGTCTGACTTGGAGGAG-3′; reverse-GACCCACATGCGCTCATTGTTG-3′), MFI1 (forward, 5′-AGATCAACACCTGTGTGGAAC-3′; reverse-CCTTACGACGTGACCAAGA-3′), Igs15 (forward, 5′-GGTGCCTGAATCTTGCAT-3′; reverse-TGGAAAGGTAAGACCGTCCT-3′), JIF7 (forward, 5′-GAACATGTCTGACTTGGAGGAG-3′; reverse-GACCCACATGCGCTCATTGTTG-3′), Oas2 (forward, 5′-AGTTCTTACTGAGGAAATG-3′; reverse-CAGACGGCTCTTGGACTCCT-3′), If44 (forward, 5′-GAACATGTCTGACTTGGAGGAG-3′; reverse-GACCCACATGCGCTCATTGTTG-3′), Oas3 (forward, 5′-TCTGTGCTGCTAAACATCAC-3′; reverse-GATGACGAGTTCGA-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-GATGACGAGTTCGA-3′), Oas3 (forward, 5′-TCTGTGCTGCTAAACATCAC-3′; reverse-GATGACGAGTTCGA-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-GATGACGAGTTCGA-3′), Oas3 (forward, 5′-TCTGTGCTGCTAAACATCAC-3′; reverse-GATGACGAGTTCGA-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-GATGACGAGTTCGA-3′), Oas3 (forward, 5′-TCTGTGCTGCTAAACATCAC-3′; reverse-GATGACGAGTTCGA-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-CGACGAGTTCGATATGTCAGC-3′; reverse-GATGACGAGTTCGA-3′). Ct values were normalized to Gapdh (forward, 5′-AATGGCTCCCTGTGATGAT-3′; reverse-CTCAGATGGCTCATTCCAC-3′ and relative expression was calculated using the 2ΔΔCt method63.
Data availability

WES and WGS sequencing data of the 76 genes described in this article have been deposited to the EGA database under the accession number EGAD0000000104859 and are accessible to all qualified researchers meeting the data access policy (EGAP00000101134) as determined by the Data Access Committee (EGAC00000110157). The 45 and Up87,88 and ASPREE89 datasets were used as reference healthy controls, accessed through the MGBR Collaborative (http://igc.garvan.org.au/mgbr/initiatives) and are accessible to all qualified researchers meeting the MGBR data access policy (https://igc.garvan.org.au/terms/mgbbr). All variants have been submitted to ClinVar (SUB4880889 and SUB5448273). The microarray data have been deposited to the GEO database under the accession number GSE126307.

Received: 3 November 2018 Accepted: 25 April 2019

Published online: 17 May 2019

References

1. Tipton, C. M. et al. Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. Nat. Immunol. 16, 755–765 (2015).
2. Degn, S. E. et al. Clonal evolution of autoreactive germinal centers. Cell 170, 913–926 e919 (2017).
3. Donner, T., Giesecke, C. & Lipsky, P. E. Mechanisms of B cell autoimmunity in SLE. Arthritis Res. Ther. 13, 243 (2011).
4. Berland, R. et al. Toll-like receptor 7-dependent loss of B cell tolerance in lupus erythematosus. J. Exp. Med. 197, 711–723 (2003).
5. Crow, J. Y. Type I interferonopathies: a novel set of inborn errors of immunity. Ann. N. Y. Acad. Sci. 1238, 91–98 (2011).
6. Deapen, D. et al. A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis Rheum. 35, 311–318 (1992).
7. Konovsky, S. V. et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat. Genet. 40, 211–216 (2008).
8. International Consortium for Systemic Lupus Erythematosus. G. et al. Genomic association scan in women with systemic lupus erythematosus. Science 326, 429–433 (2002).
9. Alarcon-Segovia, D. et al. Familial aggregation of systemic lupus erythematosus. Arthritis Rheum. 52, 1138–1147 (2005).
10. Benham, J. et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat. Genet. 47, 1457–1464 (2015).
11. Ban, T. et al. Lyn kinase suppresses the transcriptional activity of IRF5 in the TLR-MyD88 pathway to restrain the development of autoimmunity. Immunity 45, 319–333 (2016).
12. Castillejo-Lopez, C. et al. Genetic and physical interaction of the B-cell receptor and BLK in the T-cell receptor signaling pathway. PLoS One 11, e0157920 (2016).
13. Johnson, S. A. et al. Phosphorylated immunoreceptor signaling motifs (ITAMs) exhibit unique abilities to bind and activate Lyn and Syk tyrosine kinases. J. Immunol. 155, 4596–4605 (1995).
14. Hom, G. et al. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. Proc. Natl Acad. Sci. USA 92, 9562–9567 (1995).
15. Kim, J. Y. & Ozato, K. The sequestosome 1/p62 attenuates cytokine gene expression in activated macrophages by inhibiting IFN regulatory factor 8 and TNF receptor-associated factor 6/IFN-kappaB activity. J. Immunol. 182, 2131–2140 (2009).
16. Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. Nature 542, 186–190 (2017).
17. Epi, K. C. et al. De novo mutations revealed by whole-exome sequencing in unrelated samples for genome-wide association studies. Proc. Natl Acad. Sci. USA 106, 1066–1077 (2009).
18. Sanders, S. J. et al. The type I interferonopathies. Nat. Immunol. 18, 573–580 (2017).
19. Teruel, M. & Alarcon-Riquelme, M. E. The genetic basis of systemic lupus erythematosus. Curr. Rheumatol. Rep. 18, 71 (2016).
20. Lee-Kirsch, M. A. The type I Interferonopathies. Nat. Comm. 10, 2201 (2019).
21. Up Study, C. et al. Cohort profile: the 45 and up study. Int. J. Epidemiol. 37, 941–947 (2008).
22. Lacaze, P. et al. The genomic potential of the aspirin in reducing events in the elderly and statins in reducing events in the elderly studies. Intern. Med. J. 47, 461–463 (2017).
23. Liu, Y. et al. The 45 and up study: a tool for local population health and health service planning to improve integration of healthcare. Public Health Res. Pract. 26 https://doi.org/10.17061/phrp2631629 (2016).

Acknowledgements

Personnel of the Australian Cancer Research Foundation Biomolecular Resource Facility (JCSMR). Harpreet Vohra and Mick Devoy from the MCRF facility (JCSMR). Stuart Read and Nikki Ross from the Australian Phenomics Facility. FLAG-TRA6 was a gift from Robert Brink and plasmids for BANK1 were a kind gift from Casamiro Castillejo-Lopez. RACP Jacquot NHMRC Award for Excellence, Jacquot Research Entry Scholarship, and NHMRC project grants to SHJ. NHMRC Program and project grants, and Elizabeth Blackburn Fellowship to CGV. This research/project was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. This research and generation of CRISPR mice were also supported by funding of the Australian Government’s National Collaborative Research Infrastructure Strategy to the Australian Phenomics Facility and Bioplatforms Australia. We acknowledge the MGRB Collaborative (http://sgc.garvan.org.au/mgrb/initiatives) for granting us controlled access to the 45 and Up [1,2] and ASPREE [3] datasets as reference healthy controls.

Author contributions

S.J., V.A., J.E., A. Chuah, J. Capello, A. Cook, S.P., J. Cardenas, J., M.S., I.R., I.P., M.Y., K.M., J.B., C.B., G.B., L.M., P.F.C., and T.D.A., M.F. planned and performed the experiments and analyzed the data. G.W., P.G., A.E., M.I., V.T., A.L., S.A., A.K., D.F., N.S., T.A., P.G., D.M., C.L.C., E.S., P.W., M.F., T.D.A., E.C., M.C.C., V.P., and C.G.V. provided intellectual input and/or key reagents. S.J. and C.G.V. planned and wrote the paper.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10242-9.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature communications would like to thank anonymous reviewers for their contributions to the peer review of this work.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2019