

Association of the *BANK1* R61H variant with systemic lupus erythematosus in Americans of European and African ancestry

Struan FA Grant^{1,2,3}
 Michelle Petri⁴
 Jonathan P Bradfield¹
 Cecilia E Kim¹
 Erin Santa¹
 Kiran Annaiah¹
 Edward C Frackelton¹
 Joseph T Glessner¹
 F George Otieno¹
 Julie L Shaner¹
 Ryan M Smith¹
 Andrew W Eckert¹
 Rosetta M Chiavacci¹
 Marcin Imielinski¹
 Kathleen E Sullivan⁵
 Hakon Hakonarson^{1,2,3}

¹Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; ²Department of Pediatrics and Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; ³Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA; ⁴Division of Rheumatology, Johns Hopkins School of Medicine, Baltimore, MD, USA; ⁵Division of Allergy and Immunology, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

Correspondence: Struan FA Grant or Hakon Hakonarson
 Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA
 Email grants@chop.edu or hakonarson@chop.edu

Abstract: Recently an association was demonstrated between the single nucleotide polymorphism (SNP), rs10516487, within the B-cell gene *BANK1* and systemic lupus erythematosus (SLE) as a consequence of a genome wide association study of this disease in European and Argentinean populations. In a bid for replication, we examined the effects of the R61H non-synonymous variant with respect to SLE in our genotyped American cohorts of European and African ancestry. Utilizing data from our ongoing genome-wide association study in our cohort of 178 Caucasian SLE cases and 1808 Caucasian population-based controls plus 148 African American (AA) SLE cases and 1894 AA population-based controls we investigated the association of the previously described non-synonymous SNP at the *BANK1* locus with the disease in the two ethnicities separately. Using a Fisher's exact test, the minor allele frequency (MAF) of rs10516487 in the Caucasian cases was 22.6% while it was 31.2% in Caucasian controls, yielding a protective odds ratio (OR) of 0.64 (95% CI 0.49–0.85; one-sided $p = 7.07 \times 10^{-4}$). Furthermore, the MAF of rs10516487 in the AA cases was 18.7% while it was 23.3% in AA controls, yielding a protective OR of 0.75 (95% CI 0.55–1.034; one-sided $p = 0.039$). The OR of the *BANK1* variant in our study cohorts is highly comparable with that reported previously in a South American/European SLE case-control cohort (OR = 0.72). As such, R61H in the *BANK1* gene confers a similar magnitude of SLE protection, not only in European Americans, but also in African Americans.

Keywords: systemic lupus erythematosus, African Americans, European Americans, *BANK1* gene

Introduction

Systemic lupus erythematosus (SLE) is a multi-system, autoimmune inflammatory disease with complex genetic inheritance, characterized by the production of autoantibodies and complement-mediated tissue destruction.^{1–5}

SLE primarily impacts women (~9:1 ratio), particularly during child-bearing years, with an estimated prevalence of 31 per 100,000 in populations of European ancestry.^{6,7} This disorder has a strong genetic component,^{7–9} with the estimated concordance rate among monozygotic twins (~30%) being approximately ten times the rate among dizygotic twins (~3%), which is line with a high sibling relative risk ratio ($\lambda_s = 29$).^{7,8}

During the past 20 years, many linkage and candidate gene studies have been carried out in order to identify genetic variants contributing to the pathogenesis to SLE. Replicated linkages with SLE have been reported at multiple loci. For example, haplotypes carrying the HLA class II alleles *DRB1*0301* and *DRB1*1501* are strongly established as being associated with SLE.^{10–12} Replicated associations with SLE and variants in candidate genes include the HLA region,¹² *FCGR3A*,^{13,14} *FCGR2A*,¹⁵ *PDCDI*,¹⁶ and *PTPN22*.¹⁷ More recently, variants of the genes encoding interferon regulatory factor 5 (*IRF5*) and signal transducer and activator of transcription 4 (*STAT4*) have been shown to be risk factors for SLE.^{18–21}

Recently, a number of genome wide association studies have reported remarkably strong, replicable signals with complex disease, including the *KIAA0350* gene in type 1 diabetes²² and *IL23R* in inflammatory bowel disease.²³ In January 2008, through the utilization of the Illumina HumanHap300 BeadChip, Kozyrev et al²⁴ added to this repertoire by reporting a highly significant association between SLE and the *BANK1* gene [NM_017935] on chromosome 4q24, which encodes a B-cell adaptor protein. Specifically, an common coding variant, rs10516487 (R61H), was shown to confer a strong protective effect against the disease and was then replicated in the same study in separate South American and European cohorts of patients with SLE. The authors hypothesized that this variant could contribute to sustained B cell–receptor signaling and B-cell hyperactivity, a key feature of this disease.

Although these *BANK1* findings are compelling, there are continuing concerns regarding the performance of association studies in complex traits; as such independent replication efforts are now considered mandatory.²⁵ With the many errors and biases that can blight any individual study, replication by others can ensure that the original findings are robust and can also provide a more accurate estimate of the likely effect size.^{26,27}

In this study we demonstrate that the variant, rs10516487, in the *BANK1* gene confers a similar magnitude of protection of SLE as reported previously, not only in European Americans, but also in African Americans further suggesting that the B-cell pathway may be causally linked to SLE.

Results

Association between rs10516487 and SLE risk in Caucasians

In this replication attempt, we genotyped 178 SLE cases and 1808 population-based controls (all European Americans) with the Illumina HumanHap550 Genotyping BeadChip as part of our ongoing genome-wide association study of the disease. The diagnosis of SLE was based on standard criteria (see Study Subjects section).

The probe for rs10516487 is present on our 550 K BeadChip, so as a first pass we queried the data with a single test for this SNP to investigate if this marker is associated with SLE in our European American cohort. Using a Fisher’s exact test, we observed a significant protective effect of the minor A allele on the risk of SLE. The minor allele frequency (MAF) in the cases was 22.6% while it was 31.2% in controls, yielding a protective odds ratio (OR) of 0.64 (95% CI 0.49–0.85; one-sided $p = 7.07 \times 10^{-4}$) (see Table 1). This OR is very much in line with that reported previously in a South American/European SLE case-control cohort (OR = 0.72).²⁴

Association between rs10516487 and SLE risk in African Americans

We also genotyped 148 AA SLE cases and 1894 AA population-based controls with the Illumina HumanHap550 Genotyping BeadChip as part of our ongoing genome-wide association study of the disease in this ethnicity. The Fisher’s exact test also revealed a significant protective effect of the minor A allele of rs10516487 on the risk of SLE. The MAF in the cases was 18.7% while it was 23.3% in controls, yielding a protective OR of 0.75 (95% CI 0.55–1.034; one-sided $p = 0.039$) (see Table 1). This OR is again very much in line with our observation in our European American cohort and with that reported previously in other Caucasian and South American populations.²⁴

Discussion

From an interim analysis of our ongoing genome-wide association study of SLE, we have investigated variation in the *BANK1* locus previously reported to be associated with SLE in South American and European cohorts.²⁴ Consequently, we have replicated association of this gene with SLE by demonstrating its effect in Americans of both European and African ancestry. More specifically, the uncommon coding variant, rs10516487 (R61H), was shown to confer a strong protective effect against the disease with a highly comparable odds ratio (0.64 in Caucasians; 0.75 in AA) to that reported

Table 1 European American and African American systemic lupus erythematosus (SLE) case-control association study results for the *BANK1* SNP, rs10516487 (R61H)

Ethnicity	Marker	Location (Build 36)	MAF cases	MAF controls	χ^2	One-sided p-value	OR [95% CI]
European American	rs10516487	102970099	0.226	0.312	10.190	7.07×10^{-4}	0.644 [0.491–0.845]
African American	rs10516487	102970099	0.187	0.233	3.089	0.039	0.754 [0.549–1.034]

Minor allele frequencies (MAF), allelic test one-sided p values, and odds ratios (OR) with 95% confidence intervals (CI) are shown for each SNP. The ORs shown are for the minor A allele.

previously in a South American/European SLE case-control cohort (OR = 0.72).²⁴

Although the size of the cohort in the original genome wide association study was larger²⁴, the amount of testing in our cohort was very restricted to a focused effort of specifically investigating if a SNP at a single locus also yields association in the same direction as previously reported. Our cohort is sufficiently powered to ask this straight-forward validation question.

Kozyrev et al²⁴ also described two other SNPs that were associated with SLE in their cohort, rs17266594 (branch point) and rs3733197 (A383T), which were not present on our BeadChip. However, they concluded that none of these SNPs are independent of each other, as a result of the linkage disequilibrium between themselves and rs10516487. As such the association we observe with R61H may not necessarily mean it is the underlying causative variant, and could be tagging one of these two variants or another, yet to be uncovered mutation event.

As the association we observe is indeed of a very similar magnitude to that of the original report, this independent replication confirms *BANK1* as a genuine SLE susceptibility gene; in addition we have extended this observation to a population of African ancestry. As such, the “winner’s curse” that is often seen for other complex trait susceptibility genes²⁸ is not observed in this instance. What is of particular note is that the association observed in Americans of either European or African ancestry is almost identical to that of South Americans and Europeans.

Our results lend further support for the protective role of the *BANK1* gene in SLE, suggesting that interventions at the B-cell pathway level may be of value in patients who suffer from this disease. The *BANK1* gene encodes a B-cell scaffold protein with ankyrin repeats which regulates Ca²⁺ mobilization and activation.²⁹ The efficacy of B cell directed interventions supports the importance of investigating polymorphisms which could alter the B cell activation threshold.³⁰ The variants may directly dictate splicing or some other regulatory mechanism but more likely are in linkage disequilibrium with the causative variant(s).

Once our genome-wide association study is complete, we will have the opportunity to look for other variants in the genome that are associated with SLE, as a consequence of our use of a higher resolution BeadChip. In addition, we will explore the *BANK1* gene further to elucidate other potential variants that may confer genetic susceptibility to this debilitating disorder in our cohort.

Materials and Methods

Study subjects

The case-control study consisted of 178 Caucasian SLE cases, 148 African American (AA) SLE cases, 1808 population-based Caucasians controls and 1894 AA population-based controls. All subjects were biologically unrelated ethnic Caucasian or of African ancestry. Patients and controls were all from the Mid-Atlantic region of the US.

All cases fulfilled the 1982 American College of Rheumatology (ACR) criteria for the classification of SLE.³¹ All participating subjects provided informed consent for this study. The study was approved by the respective Institutional Review Boards of the participating study sites.

Genotyping

We performed high throughput genome-wide SNP genotyping using the Illumina Infinium™ II HumanHap500 BeadChip technology.^{32,33} The Infinium™ II Assay protocol enables effective multiplexing and genome-wide SNP access through a single-base extension (SBE) method with enzymatic SNP scoring. The minimal hands-on three-day Infinium™ II protocol began with 750 ng of input DNA. In day one, genomic DNA was amplified 1000–1500-fold. Day two, amplified DNA was fragmented ~300–600 bp, then precipitated and resuspended followed by hybridization onto a BeadChip. SBE utilizes a single probe sequence ~50 bp long designed to hybridize immediately adjacent to the SNP query site. Following targeted hybridization to the bead array, the arrayed SNP locus-specific primers (attached to beads) were extended with a single hapten-labeled dideoxynucleotide in the SBE reaction. The haptens were subsequently detected by a multi-layer immunohistochemical sandwich assay, as recently described.^{32,33} Our technicians followed the standard operating procedures issued by Illumina from the step that DNA is added to the single-tube whole genome amplification up to the processing of the BeadChip genotyping files. The Illumina BeadArray Reader scanned each BeadChip at two wavelengths and created an image file. As BeadChip images were collected, intensity values were determined for all instances of each bead type, and data files were created that summarized intensity values for each bead type. These files consisted of intensity data that was loaded directly into Illumina’s genotype analysis software, BeadStudio. A bead pool manifest created from the LIMS database containing all the BeadChip data was loaded into BeadStudio along with the intensity data for the samples. BeadStudio used a normalization algorithm to minimize BeadChip to BeadChip variability.

Once the normalization was complete, the clustering algorithm was run to evaluate cluster positions for each locus and assign individual genotypes. Each locus was given an overall score based on the quality of the clustering and each individual genotype call was given a GenCall score. GenCall scores provided a quality metric that ranges from 0 to 1 assigned to every genotype called. GenCall scores were then calculated using information from the clustering of the samples. The location of each genotype relative to its assigned cluster determined its GenCall score.

The resources available for this project included the Illumina technology platform itself plus nine Tecan pipetting robotic systems, eight scanners, a laboratory information management system (LIMS) and automated allele-calling software. The workflow was robotic-based for automatic sample processing and included algorithms for quality control of genotypes. The facility infrastructure had sufficient computational power and servers for data processing and storing, including a series of computers that were integrated (warehouse setting) to perform continuous datamining of all gathered and generated datasets.

Data analysis

Both the genetic matching of the cases and controls and the statistical tests for association were carried out using the software package *plink* (<http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml>). One-sided *P*-values, odds ratios and the corresponding 95% confidence intervals were calculated for the association analyses in this replication attempt.

Acknowledgments

We would like to thank Adrienne Alexander, Chioma Onyiah, Elvira Dabaghyan, Kenya Fain, Maria Garris, Wendy Glaberson, Kisha Harden, Andrew Hill, Crystal Johnson-Honesty, Lynn McCleery, Robert Skraban, Kelly Thomas and Alexandria Thomas for their expert assistance with genotyping or data collection and management. We would also like to thank Smari Kristinsson, Larus Arni Hermannsson and Asbjörn Krisbjörnsson of Raförinn ehf for their extensive software design and contribution. This research was financially supported by the Children's Hospital of Philadelphia (genotyping was funded by the Center for Applied Genomics developmental award), NIH AI051323, NIH AR 43727, a Developmental Research Award from the Cotswold Foundation and the Hopkins General Clinical Research Center (MO1-RR 00052).

Disclosures

The authors have no conflicts of interest to disclose.

References

1. Sherer Y, Gorstein A, Fritzler MJ, Shoenfeld Y. Autoantibody explosion in systemic lupus erythematosus: more than 100 different antibodies found in SLE patients. *Semin Arthritis Rheum.* 2004;34(2):501–37.
2. Sestak AL, Nath SK, Sawalha AH, Harley JB. Current status of lupus genetics. *Arthritis Res Ther.* 2007;9(3):210.
3. Alarcon-Riquelme ME. The genetics of systemic lupus erythematosus: understanding how SNPs confer disease susceptibility. *Springer Semin Immunopathol.* 2006;28(2):109–17.
4. Wakeland EK, Liu K, Graham RR, Behrens TW. Delineating the genetic basis of systemic lupus erythematosus. *Immunity.* 2001;15(3):397–408.
5. Nath SK, Kilpatrick J, Harley JB. Genetics of human systemic lupus erythematosus: the emerging picture. *Curr Opin Immunol.* 2004;16(6):794–800.
6. Danchenko N, Satia JA, Anthony MS. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus.* 2006;15(5):308–18.
7. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH, et al. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum.* 2005;52(4):1138–47.
8. Deapen D, Escalante A, Weinrib L, et al. A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum.* 1992;35(3):311–8.
9. James JA, Kaufman KM, Farris AD, Taylor-Albert E, Lehman TJ, Harley JB. An increased prevalence of Epstein-Barr virus infection in young patients suggests a possible etiology for systemic lupus erythematosus. *J Clin Invest.* 15 1997;100(12):3019–26.
10. Goldberg MA, Arnett FC, Bias WB, Shulman LE. Histocompatibility antigens in systemic lupus erythematosus. *Arthritis Rheum.* 1976;19(2):129–32.
11. Graham RR, Ortmann WA, Langefeld CD, et al. Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am J Hum Genet.* 2002;71(3):543–53.
12. Graham RR, Ortmann W, Rodine P, et al. Specific combinations of HLA-DR2 and DR3 class II haplotypes contribute graded risk for disease susceptibility and autoantibodies in human SLE. *Eur J Hum Genet.* 2007;15(8):823–30.
13. Edberg JC, Langefeld CD, Wu J, et al. Genetic linkage and association of Fcγ receptor IIIA (CD16A) on chromosome 1q23 with human systemic lupus erythematosus. *Arthritis Rheum.* 2002;46(8):2132–40.
14. Sullivan KE, Jawad AF, Piliero LM, et al. Analysis of polymorphisms affecting immune complex handling in systemic lupus erythematosus. *Rheumatology (Oxford).* 2003;42(3):446–52.
15. Duits AJ, Bootsma H, Derksen RH, et al. Skewed distribution of IgG Fc receptor IIa (CD32) polymorphism is associated with renal disease in systemic lupus erythematosus patients. *Arthritis Rheum.* 1995;38(12):1832–6.
16. Prokunina L, Castillejo-Lopez C, Oberg F, et al. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat Genet.* 2002;32(4):666–9.
17. Kyogoku C, Langefeld CD, Ortmann WA, et al. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am J Hum Genet.* 2004;75(3):504–7.
18. Sigurdsson S, Nordmark G, Goring HH, et al. Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am J Hum Genet.* 2005;76(3):528–37.
19. Graham RR, Kozyrev SV, Baechler EC, et al. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat Genet.* 2006;38(5):550–5.

20. Graham RR, Kyogoku C, Sigurdsson S, et al. Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc Natl Acad Sci U S A*. 2007;104(16):6758–63.
21. Remmers EF, Plenge RM, Lee AT, et al. 6 STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med*. 2007;357(10):977–86.
22. Hakonarson H, Grant SFA, Bradfield JP, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature*. 2007;448(7153):591–4.
23. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006;314(5804):1461–3.
24. Kozyrev SV, Abelson AK, Wojcik J, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat Genet*. 2008;40(2):211–6.
25. Patterson M, Cardon L. Replication publication. *PLoS Biol*. 2005;3(9):e327.
26. Page GP, George V, Go RC, Page PZ, Allison DB. “Are we there yet?” Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *Am J Hum Genet*. 2003;73(4):711–9.
27. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med*. 2002;4(2):45–61.
28. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet*. 2003;33(2):177–82.
29. Yokoyama K, Su IH, Tezuka T, et al. 15 BANK regulates BCR-induced calcium mobilization by promoting tyrosine phosphorylation of IP(3) receptor. *Embo J*. 2002;21(1–2):83–92.
30. Eisenberg R. SLE – Rituximab in lupus. *Arthritis Res Ther*. 2003;5(4):157–9.
31. Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum*. 1982;25(11):1271–7.
32. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet*. 2005;37(5):549–54.
33. Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL. Whole-genome genotyping with the single-base extension assay. *Nat Methods*. 2006;3(1):31–3.

