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

Struan FA Grant1,2,3
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 Hakonarson1,2,3

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

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 Argentinian 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 × 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 PDCD1,16 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 diabetes and IL23R in inflammatory bowel disease. In January 2008, through the utilization of the Illumina HumanHap300 BeadChip, Kozyrev et al 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. 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.

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
Assocation 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).

**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.

**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. 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 previously in other Caucasian and South American populations.

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

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Marker</th>
<th>Location (Build 36)</th>
<th>MAF cases</th>
<th>MAF controls</th>
<th>$\chi^2$</th>
<th>One-sided p-value</th>
<th>OR [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>European American</td>
<td>rs10516487</td>
<td>102970099</td>
<td>0.226</td>
<td>0.312</td>
<td>10.190</td>
<td>$7.07 \times 10^{-4}$</td>
<td>0.644 [0.491–0.845]</td>
</tr>
<tr>
<td>African American</td>
<td>rs10516487</td>
<td>102970099</td>
<td>0.187</td>
<td>0.233</td>
<td>3.089</td>
<td>0.039</td>
<td>0.754 [0.549–1.034]</td>
</tr>
</tbody>
</table>

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).  

Although the size of the cohort in the original genome wide association study was larger, 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 al 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 genes 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. The efficacy of B cell directed interventions supports the importance of investigating polymorphisms which could alter the B cell activation threshold. 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. 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. 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 deoxyxynucleotide in the SBE reaction. The haptenes were subsequently detected by a multi-layer immunohistochemical sandwich assay, as recently described. 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 Skranb, 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 Asbjørn Kristjórnsson of Rafórninn 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 Al051323, 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
