Genome-Wide Association Study of Generalized Vitiligo in an Isolated European Founder Population Identifies SMOC2, in Close Proximity to IDDM8

Stanca A. Birlea1,2, Katherine Gowan1, Pamela R. Fain1,3,4 and Richard A. Spritz1,4

Generalized vitiligo is a common disorder in which patchy loss of skin and hair pigmentation principally appears to result from autoimmune loss of melanocytes from affected regions. We previously characterized a unique founder population in an isolated Romanian community with elevated prevalence of generalized vitiligo and other autoimmune diseases, including autoimmune thyroid disease, rheumatoid arthritis, and type I diabetes mellitus. Here, we describe a genome-wide association study (GWAS) of generalized vitiligo in 32 distantly related affected patients from this remote village and 50 healthy controls from surrounding villages. Vitiligo was significantly associated with single-nucleotide polymorphisms (SNPs) in a 30-kb LD block on chromosome 6q27, in close vicinity to IDDM8, a linkage and association signal for type I diabetes mellitus and rheumatoid arthritis. The region of association contains only one gene, SMOC2, within which SNP rs13208776 attained genome-wide significance for association with generalized vitiligo ($P = 8.51 \times 10^{-8}$) at odds ratio $7.445$ (95% confidence interval $= 3.56-15.53$) for the high-risk allele and population attributable risk 28.00. SMOC2 encodes a modular extracellular calcium-binding glycoprotein of unknown function. Our findings indicate that SMOC2 is a risk locus for generalized vitiligo and perhaps other autoimmune diseases.

Journal of Investigative Dermatology (2010) 130, 798-803; doi:10.1038/jid.2009.347; published online 5 November 2009

INTRODUCTION

Generalized vitiligo is the most common human pigmenta-
tion disorder, affecting approximately 0.4% of European Caucasians (CEU) (Howitz et al., 1977). Generalized vitiligo is characterized by acquired, progressive, multifocal patches of white skin and overlying hair that can result in significant social stigmatization, especially in patients from darker-skinned ethnic groups. Most evidence indicates that generalized vitiligo is an organ-specific autoimmune disease directed against melanocytes (Ongenae et al., 2003; Rezaei et al., 2007), and indeed about 20% of vitiligo patients (and their close relatives) manifest concomitant occurrence of other autoimmune diseases, particularly autoimmune thyroid disease, rheumatoid arthritis, late-onset type I diabetes mellitus, psoriasis, pernicious anemia, systemic lupus erythematosus, and Addison’s disease (Alkhateeb et al., 2003).

Nonetheless can represent important candidate genes that identify potential novel pathways of disease.

Nevertheless, heritable biological properties of the melanocyte or other factors, combined with environmental triggers, may contribute to loss of immune tolerance and ultimately autoimmunity directed against melanocytes (Boissy and Spritz, 2009).

Family clusters of vitiligo cases are not uncommon, occurring in a non-Mendelian pattern suggestive of polygenic, multifactorial inheritance (Spritz, 2007, 2008). Genetic linkage and association studies have implicated a number of genes in vitiligo pathogenesis, especially genes involved in immune function (Spritz, 2007, 2008). However, these loci account for a relatively small fraction of total disease liability.

Genetically isolated “founder populations” afford special opportunities to identify genes involved in susceptibility to disease, as founder populations may have elevated prevalence of specific diseases and reduced heterogeneity of causal genetic and environmental risk factors compared with more outbred populations (Wright et al., 1999). Accordingly, susceptibility alleles that represent relatively minor genetic risk factors for complex diseases in the general population may become amplified and constitute major risk alleles in a founder population, and thus may be localized using less dense maps and smaller sample sizes than similar studies conducted in more outbred populations (Wittke-Thompson et al., 2007). Although complex disease alleles detected as major signals in a unique genetic isolate may be problematic to replicate in the wider, outbred population, they nevertheless can represent important candidate genes that identify potential novel pathways of disease.

1Human Medical Genetics Program, University of Colorado Denver, Aurora, Colorado, USA; 2Department of Dermatology, University of Colorado Denver, Aurora, Colorado, USA; 3Barbara Davis Center for Childhood Diabetes, University of Colorado Denver, Aurora, Colorado, USA; 4Department of Pediatrics, University of Colorado Denver, Aurora, Colorado, USA

Correspondence: Richard A. Spritz, Human Medical Genetics Program, University of Colorado Denver, Anschutz Medical Campus, PO Box 6511, MS 8300, Aurora, Colorado 80045, USA. E-mail: Richard.Spritz@ucdenver.edu

Abbreviations: CEU, non-Hispanic Caucasians of European origin; GWAS, genome-wide association study; SNP, single-nucleotide polymorphism

Received 28 July 2009; revised 16 September 2009; accepted 23 September 2009; published online 5 November 2009
We have studied a geographically and genetically isolated community in a remote region of northern Romania, in which the prevalence of generalized vitiligo is 2.9%, at least eight times higher than in other European Caucasian groups (Birlea et al., 2008). Similarly, the prevalence of other autoimmune diseases that are epidemiologically associated with vitiligo (autoimmune thyroid disease, rheumatoid arthritis, adult-onset type I diabetes mellitus) are also elevated in this village, both in vitiligo patients and in their first-degree relatives. This village was founded approximately 400 years ago by only three families, and all affected individuals derive from a relatively small number of common ancestors (Birlea et al., 2008); this village thus constitutes a unique founder population for generalized vitiligo.

We have taken advantage of the opportunity afforded by this special founder population to search for a generalized vitiligo susceptibility locus in this isolated village, in which reduced genetic heterogeneity enables testing a much smaller sample size than in a typical outbred population. We carried out a genome-wide association study (GWAS) of 310,598 SNPs in 32 villagers affected by generalized vitiligo from the Romanian founder population and 50 unrelated controls from surrounding villages. One SNP achieved genome-wide significance, rs13208776, located in \( \text{SMOC2} \) in distal chromosome 6q. Our results indicate that \( \text{SMOC2} \) is an important candidate susceptibility locus for generalized vitiligo in this isolated community, which may represent a microcosm of disease susceptibility in the broader CEU population.

RESULTS AND DISCUSSION

We initially analyzed 32 of the most distantly related villagers affected with generalized vitiligo, carefully selected to minimize potential false association because of genetic relatedness, and 50 unrelated controls from immediately surrounding villages. All cases met strict clinical diagnostic criteria for generalized vitiligo (Tateb and Picardo, 2007). Subject DNAs were analyzed using Illumina Infinium Human Hap300 or CNV370 BeadChip microarrays, which interrogated 310,598 SNPs in common between the two platforms. We excluded two controls because of genotyping call rates <99%. In addition, we carried out genetic matching (Luca et al., 2008) to control for population stratification by removing genetic outliers, thereby excluding four controls classified as outliers. We adjusted for relatedness among cases using CCREL software (Browning et al., 2005), which estimated that the 32 distantly related cases corresponded to an effective sample size of 30.1 independent cases. After quality control procedures, we analyzed a final dataset that included 297,342 SNPs successfully genotyped in 32 distantly related cases and 44 unrelated controls.

Following genetic matching to minimize population stratification, we compared SNP allele frequencies in cases versus controls using Fisher single-marker exact tests implemented in PLINK, version 1.05 (Purcell et al., 2007) and by allelic association tests that accounted for relatedness among cases using CCREL, version 0.3 (Browning et al., 2005). Quantile-quantile (Q-Q) analyses of the observed \(-\log_{10}(P\text{-values})\) from the Fisher exact tests (Figure 1a) and the observed \(\chi^2\)-test statistics from CCREL (Figure 1b) showed modest residual genomic inflation (Fisher’s exact test genomic inflation factor \(\lambda = 1.06\); CCREL genomic inflation factor \(\lambda = 1.10\)). Residual genomic inflation was effectively removed by adjustment of the test statistics for the corresponding inflation factor. After all corrections, we observed greater association than expected by chance (Table 1). As

![Figure 1](https://www.jidonline.org)
shown in Figure 2 and Supplementary Table S1, the most significant association signal was for SNP rs13208776 (nominal Fisher exact test \( P\)-value = 3.13 \( \times \) 10\(^{-8}\)), which surpassed a strict Bonferroni-corrected criterion for genome-wide significance \((P < 1.68 \times 10^{-7}; 0.05\) divided by 297,342 SNPs; Ioannidis et al., 2009), based on the \( P\)-values from both the Fisher exact test (adjusted \( P\) = 8.51 \( \times \) 10\(^{-8}\)) and from the CCREL likelihood ratio \( \chi^2 \) test (adjusted \( P\) = 9.71 \( \times \) 10\(^{-8}\)). The adjusted Fisher exact \( P\)-value for rs13208776 exceeded the 2.1% quantile of minimum \( P\)-values from 1,000 permutations of the phenotype labels, and was the only \( P\)-value outside the 95% confidence limits that were estimated empirically for each of the 20 top-ranking GWAS signals. Together, these analyses show that rs13208776 surpasses a conservative Bonferroni-corrected significance threshold, even after adjustment for relatedness among cases and for genomic inflation.

**Table 1. Number of significant associations identified in GWAS of Romanian population isolate for generalized vitiligo**

<table>
<thead>
<tr>
<th>Level of significance</th>
<th>Observed</th>
<th>Observed adjusted(^{1})</th>
<th>Expected</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01-0.05</td>
<td>15,998</td>
<td>13,548</td>
<td>11,894</td>
<td>1.13</td>
</tr>
<tr>
<td>0.001-0.01</td>
<td>4,536</td>
<td>3,299</td>
<td>2,676</td>
<td>1.23</td>
</tr>
<tr>
<td>0.0001-0.001</td>
<td>591</td>
<td>365</td>
<td>268</td>
<td>1.36</td>
</tr>
<tr>
<td>0.00001-0.0001</td>
<td>74</td>
<td>30</td>
<td>27</td>
<td>1.11</td>
</tr>
<tr>
<td>&lt;0.00001</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2.00</td>
</tr>
</tbody>
</table>

All \( P < 0.05\) \( 21,204 \) 17,246 14,867 1.16

Observed numbers of SNPs associated with generalized vitiligo, by level of significance, before and after genomic control adjustment for population stratification, and expected number under the null hypothesis of no association.

\(^{1}\)Adjusted for inflation of test statistic by the genomic control method (Devlin et al., 2001).

SNP rs13208776 is located on chromosome 6q27, within intron 4 of the SMOC2 gene (Figure 3). Of the 10 SNPs showing the strongest association within the SMOC2 region, most were contained in two adjacent linkage disequilibrium blocks, the first comprising seven SNPs (average \( D \) = 0.97; \( r^2 \) = 0.40), including rs13208776, and the second comprising two SNPs (\( D \) = 1.00; \( r^2 \) = 0.04); rs1402|rs214479; SNP rs2144749 had an adjusted \( P\)-value 1.75 \( \times \) 10\(^{-4}\) and is substantially correlated (\( D \) = 0.72; \( r^2 \) = 0.49) with SNP rs13208776. We identified several haplotypes in this region with \( P\)-values slightly improved over SNP rs13208776, although none were significantly better. Of 297,342 SNPs tested, the 20 SNPs with the most extreme \( P\)-values associated with vitiligo are listed in Supplementary Table S1.

SMOC2, which is the only gene in this region of 6q27, is located in close proximity to IDDM8 (http://www.t1dbase.org), a genetic locus that has both linkage and association with type I diabetes (Luo et al., 1995; Davies et al., 1996; Owerbach, 2000; Cox et al., 2001) and rheumatoid arthritis (Myerscough et al., 2000). Type I diabetes and rheumatoid arthritis are autoimmune diseases that are epidemiologically associated with generalized vitiligo, both in the outbred CEU population in general (Alkhateeb et al., 2003) and in this Romanian founder population specifically (Birlea et al., 2008). Recently, SNPs within intervening sequence 4 of SMOC2 have been genetically associated with measures of pulmonary function (Wilk et al., 2007), although none of these associations remain significant after multiple-testing correction. Our findings indicate that SMOC2 is an important novel candidate gene for susceptibility to generalized vitiligo, and perhaps to other autoimmune diseases, in the isolated Romanian founder population of this study. This finding may be difficult to verify in the broader CEU population, as genetic purification in this unique population isolate has led to a very high population attributable risk for this SMOC2 susceptibility allele that may be much smaller in the outbred CEU population, even for the same variant. SMOC2 encodes a widely expressed, SPARC (BM40)-related glycoprotein that contains two thyroglobulin type-I domains, two EF-hand
calcium-binding domains, a follistatin-like domain, and a putative signal peptide (Nishimoto et al., 2002; Vannahame et al., 2003). The specific function of the SMOC-2 protein remains unknown, though roles have been suggested in angiogenesis (Rocnik et al., 2006), cell cycle regulation (Liu et al., 2009) and mitogenesis (Liu et al., 2009). Although defective calcium transport has been reported in vitiligo melanocytes and keratinocytes (Schallreuter-Wood et al., 1996), specific involvement of SMOC-2 in calcium homeostasis remains unproven.

In the skin, SMOC-2 is mainly present in the basal levels of the epidermis, and SMOC-2-stimulated attachment of primary keratinocytes in culture (Maier et al., 2008). Together, these findings are of interest in light of the suggestion that melanocyte loss in vitiligo might result from chronic cell detachment due to defective cell adhesion (Gauthier et al., 2003). Nevertheless, it is not apparent how this etiopathological mechanism might also result in the frequent concomitant occurrence of other autoimmune diseases in vitiligo patients from this isolated community, such as autoimmune thyroid disease, type 1 diabetes, and rheumatoid arthritis.

MATERIALS AND METHODS

Subjects
The genealogy and demographic characteristics of the Romanian study community, which comprises 1673 Caucasian individuals (2004 census), have been described previously (Birlea et al., 2008). Genealogical relationships were known as early as the mid-16th century, when the community was founded by three families. The population has remained substantially isolated, with a low rate of immigration; most marriages are between neighbors and involve distant consanguinity. All study participants were examined clinically, using standard diagnostic criteria for vitiligo (Taïeb and Picardo, 2007), and were interviewed regarding ancestry.

We identified 59 living and deceased villagers with vitiligo, all of whom could be linked to one large pedigree (Birlea et al., 2008), and selected the 32 most distantly related community inhabitants with generalized vitiligo (22 females, 10 males; cousins of fourth and higher degrees) for this study. Fourteen patients had other
autoimmune disorders, including autoimmune thyroid disease (10 patients), type I diabetes (4 patients), and rheumatoid arthritis (3 patients). In addition, we selected as controls 50 unrelated adult individuals (42 female and 8 male patients), without vitiligo or other autoimmune diseases and with no known ancestral ties to the founder community, collected from immediately surrounding villages in which the prevalence of vitiligo is very low, \( \sim 0.15\% \). Informed consent was obtained from all subjects and the study was conducted according to the Declaration of Helsinki Principles. The study was approved by the Ethics Committee of Iuliu Hatieganu University of Cluj-Napoca, Romania and by COMIRB at the University of Colorado Denver.

DNA preparation, genome-wide genotyping, and data quality control

DNA was isolated from peripheral blood using the Wizard Genomic DNA Purification Kit (Promega, Madison, Wisconsin) and DNA concentrations were quantified with PicoGreen (Molecular Probes, Eugene, Oregon). Genome-wide genotyping was carried out at Decode Genetics (Reykjavik, Iceland) using 3 \( \mu \)g DNA and Illumina Infinium Human Hap300 or CNV370 microarrays, with 310,598 SNPs common to both platforms and hence considered here. Genotypes were determined using Illumina Bead Studio software.

SNP genotype quality control analyses were carried out using the PLINK tool set (version 1.05) (Purcell et al., 2007). A total 13,256 SNPs were excluded because of call rates \( \leq 0.95 \) (\( n = 483 \)), minor allele frequency \( \leq 0.05 \) (Liu et al., 2008) (\( n = 12,761 \)), SNPs that mapped to both X and Y chromosomes (\( n = 2 \)), and SNPs that deviated significantly (\( P < 10^{-4} \)) from Hardy–Weinberg in the control population (\( n = 10 \)). Two controls were excluded because of overall SNP call rate \( < 99\% \). Genetic matching and outlier exclusion was carried out using principal component analysis implemented in genetic matching software (Luca et al., 2008) to a correlation matrix of 13,762 independent SNPs (\( r^2 < 0.09 \)) identified using the “indep-pairwise” pruning option in PLINK. Four control individuals with ancestry coefficients exceeding six SD were excluded.

Statistical analyses

Fisher exact tests of single marker and haplotype association were carried out using PLINK, version 1.05 (Purcell et al., 2007); allelic association tests that accounted for relatedness among cases were carried out using CCREL, version 0.3 (Browning et al., 2005). Haplotype frequencies were estimated using PHASE, version 2.1 (Stephens and Donnelly, 2003), and LD blocks were delineated using HAPLOVIEW, version 3.32 (Barrett et al., 2005). A quantile–quantile (Q-Q) plot of observed and expected \(-\log_{10}(P\text{-values})\) from Fisher exact tests (Figure 1a), with expected values derived as \( P = r/(n + 1) \), in which \( r \) corresponds to the rank of \( n = 297,342 \) ordered \( P\text{-values} \), was generated and the inflation of the observed \(-\log_{10}(P\text{-values})\) was measured as the slope of the best-fitting line using R software, version 2.9.0 (http://www.r-project.org/). 95% confidence limits for the 20 top-ranked \(-\log_{10}(P\text{-values})\) on the plot were estimated from 1,000 permutations of the phenotype labels. A Q-Q plot corresponding to the observed versus expected \( \chi^2 \)-statistics from the CCREL analysis (Figure 1b) was generated and the genomic inflation factor for the observed \( \chi^2 \)-statistics was estimated as the observed median \( \chi^2 \) divided by the expected median \( \chi^2 \) (Devlin et al., 2001).

CONFlict of interest

The authors state no conflict of interest.

Acknowledgments

We thank the villagers for their cooperation, and Drs Augustin Pop, Marinela Micle, and Florina Barbir, Maria Barbos, and Adrian Cucuian for their help in collecting samples and interviewing villagers. This work was supported by a grant from the American Skin Association and by grant AR45584 from the National Institutes of Health.

Supplementary material

Supplementary material is linked to the online version of the paper at http://www.nature.com/jid

References

Myerscough A, John S, Barrett JH et al. (2000) Linkage of rheumatoid arthritis to insulin-dependent diabetes mellitus loci: evidence supporting a
hypothesis for the existence of common autoimmune susceptibility loci. *Arthritis Rheum* 43:2771–5


Purcell S, Neale B, Todd-Brown K et al. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81:559–75


