ARTICLE

Family Clustering of Autoimmune Vitiligo Results Principally from Polygenic Inheritance of Common Risk Alleles

Genevieve H.L. Roberts,[1] Subrata Paul,[2] Daniel Yorgov,[3] Stephanie A. Santorico,[1,2,4,5] and Richard A. Spritz[1,6,*]

Vitiligo is an autoimmune disease that results in patches of depigmented skin and hair. Previous genome-wide association studies (GWASs) of vitiligo have identified 50 susceptibility loci. Variants at the associated loci are generally common and have individually small effects on risk. Most vitiligo cases are "simplex," where there is no family history of vitiligo, though occasional family clustering of vitiligo occurs, and some "multiplex" families report numerous close affected relatives. Here, we investigate whether simplex and multiplex vitiligo comprise different disease subtypes with different underlying genetic etiologies. We developed and compared the performance of several different vitiligo polygenic risk scores derived from GWAS data. By using the best-performing risk score, we find increased polygenic burden of risk alleles identified by GWAS in multiplex vitiligo cases relative to simplex cases. We additionally find evidence of polygenic transmission of common, low-effect-size risk alleles within multiplex-vitiligo-affected families. Our findings strongly suggest that family clustering of vitiligo involves a high burden of the same common, low-effect-size variants that are relevant in simplex cases. We furthermore find that a variant within the major histocompatibility complex (MHC) class II region contributes disproportionately more to risk in multiplex vitiligo cases than in simplex cases, supporting a special role for adaptive immune triggering in the etiology of multiplex cases. We suggest that genetic risk scores can be a useful tool in analyzing the genetic architecture of clinical disease subtypes and identifying subjects with unusual etiologies for further investigation.

Introduction

Vitiligo is an autoimmune disease in which destruction of skin melanocytes leads to patches of depigmented skin and hair. Although most vitiligo cases are "simplex," wherein there is no family history of vitiligo, about 9% of cases are "multiplex," belonging to families with more than one close affected relative. Inheritance of vitiligo within multiplex-affected families typically appears non-Mendelian, and early segregation analyses of such families suggested additive, polygenic inheritance with heritability between 50% and 75%.[1–3]

A polygenic nature of vitiligo is also supported by the results of subsequent genome-wide association studies (GWASs) in European-derived whites; these studies have identified 50 contributory loci, and each locus contributes a small amount to overall vitiligo heritability.[4–6] Like in most other genetically complex diseases, most vitiligo GWAS loci involve common, low-to-moderate effect-size variants (Figure 1). And, like in other autoimmune diseases, the corresponding identified genes involve the regulation of immune cells and apoptosis, as well as melanocyte components that can act as autoantigens.

We have assembled a highly curated collection of over 400 multiplex-vitiligo-affected families and over 4,000 simplex-vitiligo-affected subjects of European descent. On the basis of loci identified in previous GWASs, we investigated whether the genetic risk factors for multiplex and simplex vitiligo are fundamentally similar or different. We considered two alternative hypothetical models of genetic architecture in multiplex-vitiligo-affected families versus simplex-affected subjects. First, we considered that in multiplex-affected families, rare, high-penetrance alleles might be the major driver of vitiligo risk; such alleles would have little impact on risk for simplex vitiligo cases and would be difficult to identify by GWAS. This hypothesis was supported by the identification of several loci in previous genetic linkage analysis of multiplex-vitiligo-affected families; none of these loci overlap any vitiligo GWAS loci.[7,8] Alternatively, we considered that multiplex-vitiligo-affected families might segregate a high burden of the same common, low-to-moderate effect-size risk alleles detected by vitiligo GWAS. This hypothesis was supported by previous segregation analyses of multiplex-affected families that suggested polygenic inheritance.[3,9,10] To explore these alternative possibilities, we developed several different polygenic risk scores based on data from previous vitiligo GWASs. We then used the best-performing risk score to compare polygenic risk in simplex-affected subjects versus in the probands of multiplex-affected families. Our results suggest that multiplex-vitiligo-affected families do in fact segregate a high burden of common, low-to-moderate effect-size risk alleles identified by GWAS. Thus, genetic risk of vitiligo in both
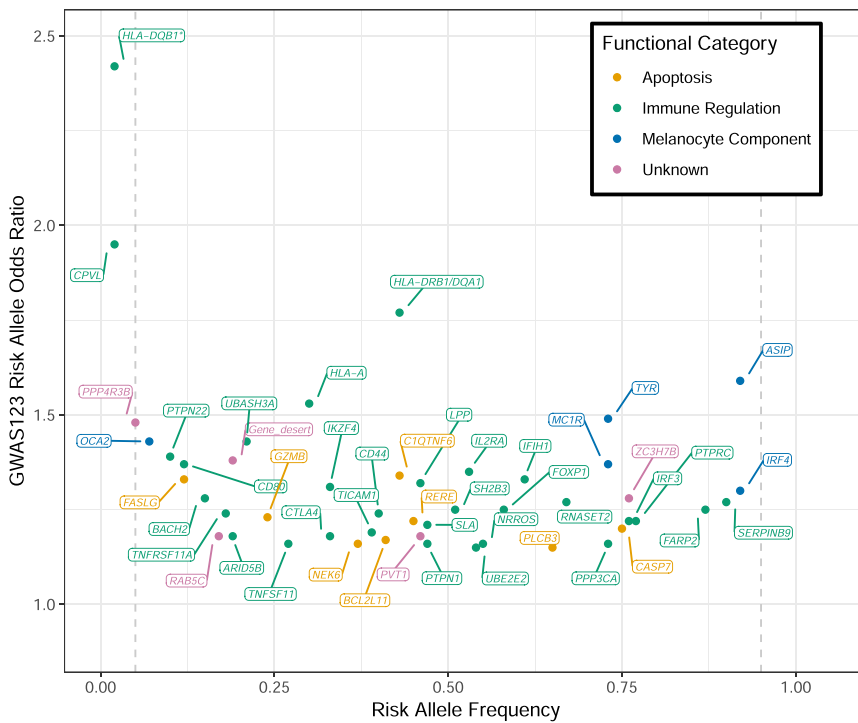
Figure 1. Distribution of Risk Allele Frequency and Estimated Odds Ratio for 48 Autosomal Variants Previously Identified in Vitiligo GWAS123

Each dot represents the most-associated variant for a vitiligo susceptibility locus, and each locus is labeled with a gene in close proximity to the most-associated variant in the locus. Dot and label colors represent the likely functional category of the locus, designated by manual review of each locus: orange = apoptosis, green = immune regulation, blue = functional component of the melanocyte, and pink = unknown function. * denotes an *HLA-DQB1* locus specifically associated in early-onset vitiligo cases;[12] the effect size used for the *HLA-DQB1* variant here was derived from all vitiligo cases, regardless of the age of onset.

multiplex-affected families and simplex-affected subjects appears to be generally similar, largely involving the same common risk alleles and thus involving the same underlying pathobiological processes.

Subjects and Methods

Study Subjects

All subjects were from North America and Europe and were of self-described non-Hispanic/Latino European-derived white ancestry. All cases met diagnostic criteria for generalized vitiligo.[11] We obtained written informed consent from all participants, and the study was approved by the institutional review board at each participating center with oversight from the Colorado Multiple Institutional Review Board (COMIRB).

Subjects consisted of 4,668 unrelated vitiligo-affected individuals and 39,436 controls from our three previous vitiligo GWASs (referred to in sum as GWAS123): GWAS1,[4] $n_{cases} = 1,377$, $n_{controls} = 14,401$; GWAS2,[5] $n_{cases} = 412$, $n_{controls} = 5,189$; GWAS3,[6] $n_{cases} = 1,052$, $n_{controls} = 17,665$), as well as an independent replication study[6] ($n_{cases} = 1,827$, $n_{controls} = 2,181$). Subjects in the replication study cohort were genotyped for the most-associated variant for all genome-wide significant loci (Table S1).

Genotyping and Quality Control

Some individuals in GWAS123 were excluded on the basis of pairwise identity-by-descent sharing $(\hat{\pi}) \geq 0.0884$; affected subjects were retained over controls, and individuals with the highest SNP call rate were retained otherwise; identity-by-descent analysis was performed with KING[12] version 1.4. For each GWAS, variants with minor allele frequency (MAF) > 0.01, Hardy-Weinberg equilibrium $P > 10^{-4}$, and INFO scores > 0.5 were retained. A total of 8,615,281, 8,632,074, and 8,616,642 autosomal variants passed the QC process for GWAS1, 2, and 3, respectively.

Vitiligo Multiplex-Affected Family Cohort

Multiplex-affected families were identified by manual, post hoc review of all European-derived vitiligo-affected subjects and relatives in our collection. When a single multiplex "proband" was included in a prior GWAS or replication study, this subject was virtually always a "true" proband (the initial contact subject from the family that enrolled in the study).

Diagnoses of vitiligo in multiplex-affected subjects and reportedly unaffected family members were verified by manual review of all available phenotype information for each subject. Subjects with apparent segmental vitiligo or with an uncertain diagnosis were assigned an unknown phenotype ($n_{unknown} = 83$). In total, we identified 444 unique multiplex-affected families with at least two relatives with generalized vitiligo and for which at least one subject had a DNA sample. Across all 444 multiplex-affected families, we identified a total 1,332 subjects with generalized vitiligo and 6,476 relatives who reported being unaffected with vitiligo.

After aligning all available genotype data from the previous GWASs and replication studies with pedigree data, we identified 324 unrelated multiplex-affected family probands with genome-wide data from GWAS123, plus an additional 78 unrelated multiplex-affected probands genotyped at all confirmed vitiligo GWAS loci in our replication study. Thus, 8.6% (402/4,668) of the combined GWAS123 and replication study vitiligo-affected subjects are multiplex probands.

Extended Multiplex-Affected Family Analysis

In our previous vitiligo GWASs and replication studies, only a single proband from each multiplex-affected family was genotyped. However, in a previous family-based association study,[4] we genotyped tagSNPs for 14 of the 50 confirmed vitiligo GWAS loci (Table S2) in all available members of 330 multiplex-vitiligo-affected families. This family-based study thus provided limited relevant genotype information for extended families of multiplex-affected probands. We used these data to investigate the transmission of polygenic risk within families. From the total set of 330 multiplex-affected families, we retained trios that consisted of a vitiligo-affected proband and two unaffected parents for whom genotype data were available for all trio members. We

additionally removed subjects with >1 missing genotype among the 14 loci. This resulted in 107 affected subjects and 148 unaffected parents from 71 multiplex-affected families whom we used to assess polygenic transmission with the 14-SNP "FAMILY" risk score described below.

Calculation of Ancestry-Based Principal Components

Genetic sub-structure for each of the three GWAS cohorts was determined by Spectral-GEM.[14] We used best-guess genotypes if the genotype posterior probability was ≥ 0.9; otherwise, we set genotypes to missing. Variants with call rate ≤ 0.99 were excluded. We performed linkage disequilibrium (LD) pruning by using PLINK v.1.07, resulting in 113,595 variants ($r^2 < 0.17$ in sliding windows of 500 variants) after excluding four genomic regions (MHC, *LCT*, chromosome 8, and chromosome 11 inversion regions). Sparser sets of markers were created by random sampling from the set of 113,595 variants, yielding additional datasets with 15,000, 40,000, 65,000, and 90,000 variants. We performed spectral-GEM cluster analyses separately for GWAS1, GWAS2, and GWAS3, and we used the pseudo-F statistics computed from the GEM-derived clusters to determine the variant set that best describes genetic substructure (90,000 variants). Spectral-GEM principal components (PCs) with logistic regression family-wise error rate < 0.1 were included as covariates in case-control analyses (GWAS1: PC2, PC8; GWAS2: PC1, PC2, PC3, PC4; and GWAS3: PC1, PC3, PC4, PC5, PC7, PC10, PC12, and PC14).

The vitiligo-affected subjects were relatively homogeneous with respect to European ancestry (Figure S1); however, we generated affected-only PCs by combining affected subjects from GWAS123 with 1000 Genomes Project (1KGP) European-derived (EUR) subjects at all variants genotyped in both groups, excluding ambiguous (A/T or G/C) SNPs. Variants with an overall missing genotype rate >0.05 in the combined set of subjects were removed. Variants were then LD-pruned with PLINK-1.9 command–indep-pairwise 50 5 0.2. Some 1KGP subjects were excluded on the basis of pairwise identity-by-descent sharing ($\hat{\pi}$) ≥ 0.0884, resulting in a total of 444 EUR 1KGP subjects. The top ten PCs were calculated from these EUR 1KGP individuals, and the vitiligo-affected subjects from GWAS123 were then projected onto the principal component analysis space. Only PC3 was significantly associated ($p < 0.05$) with the CONFIRMED risk score (described below) in affected subjects; accordingly, PC3 was included in affected-only analyses of the CONFIRMED risk score.

Construction and Comparison of Vitiligo Polygenic Risk Scores

We used 10-fold cross-validation to estimate risk score functions and to compare the performance of polygenic risk scores by using four different approaches in the GWAS123 affected-control cohorts described above. Cross-validation sets were defined by randomly assigning individuals in each GWAS cohort to one of ten groups of approximately equal size. Within each GWAS, a training set, i, was formed by leaving out a cross-validation set. We then performed vitiligo GWAS for each training set by logistic regression using dosage values as genotypes and including as covariates the ancestry-based PCs described above. We the combined effect estimates over GWAS1, GWAS2, and GWAS3 by using inverse variance weighting,[15] generating meta-analysis summary statistics for each of the ten training sets.

A risk score function was defined based on the variant effect estimates derived from the meta-analysis summary statistics from each training set as $S_{ij} = \hat{\beta}_i' G_{ij}$, where G_{ij} is the vector of best-guess genotypes for variants selected for inclusion in the score function, $\hat{\beta}_i$ is the vector of variant effect sizes from training set i, and j indexes an individual. Best-guess genotypes were called if a genotype had posterior probability of at least 0.499 and otherwise were set to missing. After hard-calls, 89,158, 81,217, and 84,031 SNPs were excluded on the basis of $>2\%$ missingness for GWAS1, 2, and 3, respectively. When genotype data were missing, we used the expected genotype value based on allele frequency, $2p^2 + 2p(1-p)$, where p is the risk allele frequency for the variant with missing data, estimated in both affected subjects and controls for each GWAS study cohort separately.

By using the test set left out of cross-validation set i, we estimated the probability of disease for each individual by logistic regression including the risk score and ancestry-based PCs. The area under the curve (AUC) for the receiver operator characteristic (ROC) curve was derived based on the probability of disease. We calculated AUC, with standard error, for each score function for each GWAS by using the pROC package version 1.14.0 in R version 3.6.0. To combine AUC over GWAS1, GWAS2, and GWAS3, we used inverse-variance weighting, resulting in an AUC for each cross-validation set.

We compared AUC for polygenic risk scores by four different approaches: (1) ALL, (2) CLUMPED, (3) CONFIRMED, and (4) FAMILY. For risk scores ALL and CLUMPED, variants were filtered to those falling below different p value thresholds (α_T), and AUC was compared at each threshold; p value thresholds considered were 5×10^{-8}, 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1. For the risk scores denoted "ALL," all variants below the p value threshold were included. For the risk scores denoted "CLUMPED," we reduced the set of variants considered by LD clumping, implemented in PLINK. Using a greedy algorithm, clumps start with a lead variant defined by the smallest p value and other variants are assigned to a clump on the basis of $r^2 > 0.2$ within 250 KB of the lead variant. For the risk score denoted "CONFIRMED," only the 48 autosomal variants that exceeded genome-wide significance ($p < 5 \times 10^{-8}$) in previous vitiligo GWASs and confirmed by independent replication[6] were included (Table S1). The two MHC Class II variants (dbSNP: rs145954018 and dbSNP: rs9271597) included in the CONFIRMED risk score are correlated;[13] we therefore applied a special coding strategy for these two variants: if 1 or 2 copies of the dbSNP: rs145954018 risk allele were present, we used $1 \times \hat{\beta}_{rs145954018}$ (reflecting a dominant effect), but if 0 copies of the dbSNP: rs145954018 risk allele were present, we used $G_{rs9271597,j} \times \hat{\beta}_{rs9271597}$.

For the risk score denoted "FAMILY," we used a set of 14 variants (Table S2) identified in GWAS1.[4] The 14 variants in the FAMILY risk score represent those with the highest MAFs and effect sizes identified by vitiligo GWASs and included in the CONFIRMED risk score.

To assess the extent of upward bias in CONFIRMED risk score effect sizes estimated from the GWAS123 variant-discovery cohort, we fit a model assuming normal errors with

$$E(Y_{CONFIRMED}) = \beta_0 + \beta_{Vitiligo} X_{Vitiligo} + \beta_{Rep} X_{Rep} + \beta_{Vitiligo*Rep} X_{Vitiligo} X_{Rep}$$

where $Y_{CONFIRMED}$ represents the normalized CONFIRMED risk score (mean centered and scaled on the basis of GWAS123 controls), β_0 represents the reference group of GWAS123 controls, $X_{Vitiligo}$ is 1 if the subject is a vitiligo-affected individual and 0 if the subject is a control, X_{Rep} is 1 if the subject was genotyped in the replication study and
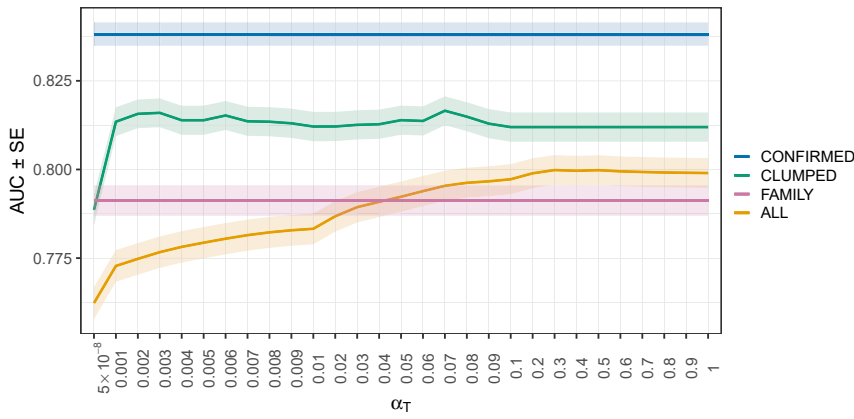
Figure 2. Comparison of AUC for Four Different Approaches for Vitiligo Risk Score Calculation at Multiple P Value Thresholds ($\propto_T$)

Four different approaches are shown: ALL (orange), CLUMPED (green), CONFIRMED (blue), and FAMILY (pink). The solid line represents the mean AUC across the 10-fold cross-validation sets, and the shaded region represents the standard error of the mean AUC.

0 if the subject was genotyped in GWAS123, and $\beta_{Vitiligo*Rep}$ represents an interaction term. To test the hypothesis that $\beta_{Vitiligo*Rep} = 0$, we used a two-sided Wald test and considered $p < 0.05$ to represent a significant difference in the relationship between the CONFIRMED risk score and vitiligo affliction in GWAS123 versus in the independent replication study.

We estimated odds ratios (ORs) of developing vitiligo for subjects belonging to the upper tails of the CONFIRMED and FAMILY risk score distributions (i.e., above the 80th, 90th, 95th, or 99th percentile of risk score among controls) by fitting a logistic regression model:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_{High-Risk\ Group}X_{High-Risk\ Group} + \sum_{i=1}^{14}\beta_{PC_i}X_{PC_i}$$

where $\pi = P(Vitiligo)$ and $X_{High-Risk\ Group}$ was coded as 1 if the subject belonged to the high risk group and 0 otherwise. X_{PC_i} corresponds to the Spectral-GEM PCs described above. PCs were not included for replication study subjects because genome-wide genotype data were not available.

Affected-Only Comparisons of the CONFIRMED Risk Score

We calculated risk scores by using the best-performing score, the CONFIRMED risk score (denoted S_C), in all vitiligo cases. For CONFIRMED risk score calculation, we used the variants and effect sizes, defined in Table S1, from the published GWAS123 meta-analysis.[6] We compared the CONFIRMED risk score between multiplex-affected probands and simplex-vitiligo-affected subjects by linear regression, assuming normal errors, with

$$E(Y_{CONFIRMED}) = \beta_0 + \beta_{multiplex}X_{multiplex} + \beta_{PC3}X_{PC3}$$

where $Y_{CONFIRMED}$ is the normalized (mean centered and scaled in GWAS123 controls) CONFIRMED risk score, β_0 represents the reference simplex-affected group, $X_{multiplex}$ is coded 0 if the case is simplex and 1 if the case is multiplex, and X_{PC3} represents ancestry-based PC3 described above. To test the hypothesis that $\beta_{multiplex} = 0$, we used a two-sided Wald test, where $p < 0.05$ represented a significant difference in CONFIRMED risk score between simplex and multiplex cases.

We categorized GWAS probands into five different groups by total number of affected relatives in the family: simplex, 2 affected, 3 affected, 4 affected, and ≥ 5 affected. We fit a linear regression model, assuming normal errors:

$$E(Y_{CONFIRMED}) = \beta_0 + \beta_{Family\_Type}X_{Family\_Type} + \beta_{PC3}X_{PC3}$$

where X_{Family_Type} was coded in an ordinal fashion (simplex = 0; 2 affected = 1; 3 affected = 2; 4 affected = 3; ≥ 5 affected = 4). We used a two-sided Wald test to test the hypothesis $\beta_{Family\_Type} = 0$, where $p < 0.05$ represented a significant relationship between the polygenic risk score and the number of affected relatives.

We analyzed the transmission of polygenic risk within multiplex-affected families by using the polygenic transmission disequilibrium test (pTDT)[16] in the Extended Multiplex-Affected Family cohort described above. We used the FAMILY risk score because genotypes were only available for 14 of the confirmed loci in probands' relatives. We considered pTDT $p < 0.05$ to represent significant average deviation between the unaffected, mid-parent risk score and affected offspring's risk score.

We compared the single variant components of the CONFIRMED risk score in multiplex-affected probands versus simplex-vitiligo-affected subjects. Given an individual's genotype, G_k, at variant k and the effect size, $\widehat{\beta}_k$, we calculated the proportion of their CONFIRMED risk score explained as $Y_{Proportion\ CONFIRMED\ Explained} = \widehat{\beta}_k G_k/S_C$. We then compared the proportion of the CONFIRMED risk score explained by each of the 48 variants in simplex- and multiplex-vitiligo-affected probands by fitting a linear regression with normal errors:

$$E(Y_{Proportion\ CONFIRMED\ Explained}) = \beta_0 + \beta_{multiplex}X_{multiplex} + \beta_{PC3}X_{PC3}.$$

To test the hypothesis that $\beta_{multiplex} = 0$, we used a two-sided Wald test and controlled false discovery to a rate of 5%.

For individual variants that explained significantly different proportions of the CONFIRMED risk score in multiplex- versus simplex-affected probands, we assessed the allele frequency in multiplex- versus simplex-affected probands by fitting a logistic regression model:

$$\log\left(\frac{\varphi}{1-\varphi}\right) = \beta_0 + \beta_{rs}X_{rs} + \beta_{PC3}X_{PC3}$$

where $\varphi = P(Multiplex)$ and X_{rs} represents the risk allele counts (0, 1, or 2) of the variant. We used a two-sided Wald test to test the hypothesis $\beta_{rs} = 0$, where $p < 0.05$ represented a significant difference in allele frequency in multiplex and simplex cases.

Results

Selection of the Best-Performing Polygenic Risk Score

Figure 2 shows the mean AUC from 10-fold cross-validation as a function of the p value threshold for inclusion of a variant in each polygenic risk score. Results are given for each of the four approaches that we used for calculating

Table 1.   Risk of Vitiligo Associated with High CONFIRMED and FAMILY Risk Scores

High-Risk Score Group[a]	Reference Group	CONFIRMED Risk Score			FAMILY Risk Score[c]	
		Minimum Risk Score[b]	GWAS123 OR (95% CI)	Replication OR (95% CI)	Minimum Risk Score[b]	GWAS123 OR (95% CI)
Top 20% of distribution	Remaining 80%	0.84	5.32 (4.90–5.77)	4.87 (4.21–5.64)	0.84	3.20 (2.95–3.47)
Top 10% of distribution	Remaining 90%	1.29	6.21 (5.70–6.76)	5.26 (4.42–6.29)	1.30	3.50 (3.19–8.84)
Top 5% of distribution	Remaining 95%	1.66	7.33 (6.64–8.08)	6.20 (4.96–7.79)	1.69	3.83 (3.42–4.29)
Top 1% of distribution	Remaining 99%	2.35	11.10 (9.41–13.10)	8.79 (5.85–13.78)	2.39	4.51 (3.63–5.59)

Abbreviations are OR - odds ratio and CI = confidence interval.
[a]Risk score distribution defined by GWAS123 controls.
[b]Minimum normalized risk score to be assigned to the high-risk group; the normalized score represents the standard deviations above the mean in GWAS123 controls.
[c]FAMILY risk score cannot be calculated in the replication study because replication samples were not genotyped at all of the variants included in the FAMILY risk score.

risk score (ALL, CLUMPED, CONFIRMED, and FAMILY). AUC varied from 0.76 (SE = 0.0046) to 0.84 (SE = 0.0033) over the risk scores considered. The CONFIRMED risk score produced the highest AUC.

As shown in Figure S2, the CONFIRMED risk score distribution in affected subjects and controls appears similar in GWAS123 and in the independent replication study; however, there is significant (p = 0.013, $\widehat{\beta}_{Vitiligo*Rep}$ = -0.096 SD) inflation of the estimated association between risk score and vitiligo in GWAS123 versus the replication study. Such inflation is expected because the variants and effect sizes we used in calculating the CONFIRMED score were derived from GWAS123. Thus, the ORs comparing the upper tails of the CONFIRMED risk score to the rest of the distribution might be more accurately represented by the independent replication study.

The CONFIRMED risk score performs better in predicting vitiligo risk than does vitiligo family history. If an individual has a sibling with vitiligo, the empirical risk of that individual developing vitiligo is 0.061,[3] and the corresponding AUC[17] based on sibling family history is 0.53. Thus, the CONFIRMED risk score (AUC 0.84) substantially outperforms empirical risk prediction based on family history alone.

Vitiligo risk is extremely high among subjects with CONFIRMED risk score above the 80th percentile (Table 1). Indeed, the risk of developing vitiligo among individuals in the top percentile of the CONFIRMED risk score (replication OR_{99} = 8.79) is substantially higher than for many other genetically complex diseases, including coronary artery disease (OR_{99} = 4.83), atrial fibrillation (OR_{99} = 4.63), type II diabetes (OR_{99} = 3.30), inflammatory bowel disease (OR_{99} = 3.87), and breast cancer (OR_{99} = 3.36).[18]

Comparison of Polygenic Risk Score in Simplex- and Multiplex-Vitiligo-Affected Probands

To assess whether the elevated burden of common variants identified by GWAS contributes to family clustering of viti-

ligo cases, we first compared the CONFIRMED risk score in 391 multiplex- and 4,102 simplex-vitiligo-affected probands. As shown in Figure 3, the CONFIRMED risk score was significantly higher (p = 2.62 × 10^{-4}, $\widehat{\beta}_{multiplex}$ = 0.21 SD) in multiplex-affected probands than in simplex-affected subjects, suggesting that the common risk variants identified by vitiligo GWASs represent an important component of genetic risk in multiplex-affected families.

To determine how the burden of common vitiligo risk alleles relates to the number of affected relatives in multiplex-affected families, we assessed the relationship between a proband's CONFIRMED risk score and the total number of affected relatives in their family. As shown in Figure 4, we found a significant positive relationship between increasing risk score and increasing number of affected relatives (p = 1.85 × 10^{-4}, $\widehat{\beta}_{Family\_Type}$ = 0.10 SD per number of affected relatives). This suggests that the burden of common vitiligo risk alleles is an important underlying cause of family clustering, including in families with many affected relatives.

To confirm polygenic inheritance of common, low-to-moderate effect size loci in vitiligo family clustering, we performed pTDT by using the FAMILY risk score in 71 multiplex-affected families. We found that higher polygenic risk as defined by the FAMILY risk score is significantly over-transmitted from unaffected parents to affected offspring (p = 0.011, mean = 0.25 SD, SE = 0.087 SD). Altogether, these findings strongly support a polygenic inheritance model in which a high burden of common, low-to-moderate effect risk alleles constitutes a major driver of vitiligo risk in multiplex-affected families.

Polygenic Risk Score is High in a Multiplex-Vitiligo-Affected Family Segregating a Rare, High-Penetrance Variant

The largest vitiligo multiplex-affected family in our collection includes 13 close relatives affected by vitiligo. In this large family, we previously demonstrated genetic
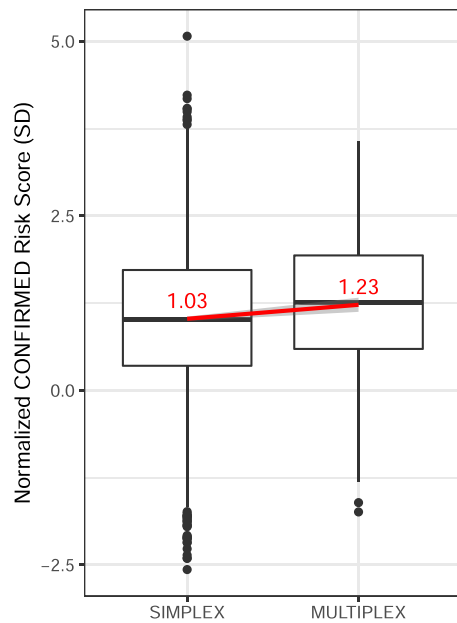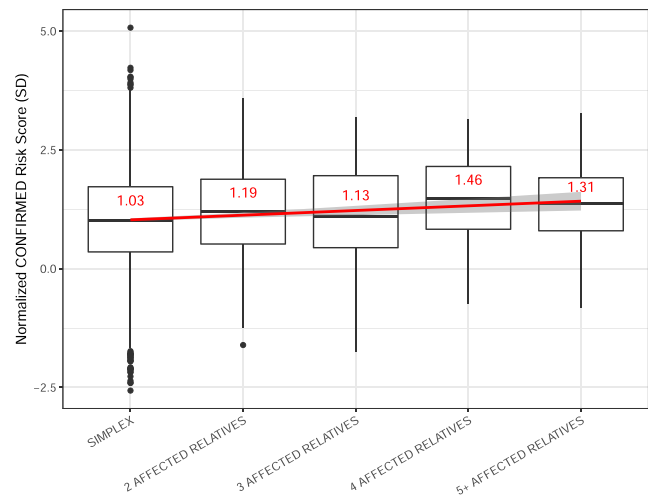
Figure 3. Comparison of CONFIRMED Risk Score in Simplex-Affected Subjects and Multiplex-Vitiligo-Affected Probands in GWAS123 + Replication Data
The y axis represents the normalized CONFIRMED risk score; units represent standard deviations (SDs) difference from the mean risk score in controls. Center horizontal lines denote the medians, and the value of the mean is shown in red text. The red line shows the fit linear regression line, and the corresponding 95% CI surrounds the line in gray. Boxes denote first through third quartiles. Each vertical bar extends from the box to the largest or smallest value, no farther than 1.5 times the inter-quartile range. Data beyond the vertical bars are considered outliers and are plotted individually.

linkage of vitiligo to chromosome 1p31.3[7] and discovered a rare functional variant (dbSNP: rs41285370G>T, $MAF_{gnomAD} = 6.4 \times 10^{-4}$) in the promoter region of *FOXD3*, which encodes a master regulator of melanocyte development.[19] The rare T allele of dbSNP: rs41285370 is present in all 13 of the vitiligo-affected subjects, and the LOD score maximizes at a penetrance of 52%.[7] Integrating the family pedigree with each individual's FAMILY risk score and dbSNP: rs41285370 genotype (Figure 5), most affected family members have high polygenic risk scores. Most affected subjects with available genotype data (8/11) fell in the 80th percentile of the FAMILY risk score or higher and thus have polygenic risk corresponding to an OR of at least 3.2 (Table 1). Altogether, our findings in this family suggest that the effect of the rare *FOXD3* promoter variant combines with the effects of common, low-penetrance risk alleles to drive especially high genetic risk of vitiligo in this remarkable family cluster.

Individual Variant Contributions to the CONFIRMED Risk Score in Multiplex versus Simplex Cases
To determine whether the increased CONFIRMED risk score in multiplex cases is driven by a particular subset of variants, we assessed each variant's average contribution to the score (Figure S3 and Table S3). Two variants, dbSNP:


Figure 4. Comparison of CONFIRMED Risk Score in Vitiligo Cases Categorized by the Number of Affected Relatives Reported by the Proband
The y axis represents the normalized CONFIRMED risk score; units represent standard deviations (SDs) difference from the mean risk score in controls. Horizontal lines denote the median, and the value of the mean is shown in red text. The red line shows the fit linear regression line, and the corresponding 95% CI surrounds the line in gray. Boxes denote first through third quartiles. Each vertical bar extends from the box to the largest or smallest value, no farther than 1.5 times the inter-quartile range. Data beyond the vertical bars are considered outliers and are plotted individually.

rs9271597 (FDR-adjusted p = 0.0046) and dbSNP: rs6059655 (FDR-adjusted p = 0.020), contributed significantly different proportions to the CONFIRMED risk score in multiplex versus simplex cases. dbSNP: rs9271597, which tags the MHC class II *HLA-DRB1/DQA1* locus, contributed an average 4.8% in multiplex cases versus 4.1% in simplex cases. Furthermore, the frequency of the dbSNP: rs9271597 risk allele is significantly higher ($p = 2.07 \times 10^{-5}$, $\hat{\beta}_{rs9271597} = 0.29$) in multiplex versus simplex cases. The dbSNP: rs6059655/*ASIP* locus contributed an average 7.5% to the risk score in multiplex cases versus 7.8% in simplex cases; however, there is no significant difference ($p = 0.20$, $\hat{\beta}_{6059655} = -0.21$) in the allele frequency of dbSNP: rs6059655 in multiplex versus simplex cases. On the basis of the allele frequency differences, we conclude that elevated contribution of dbSNP: rs9271597 to the score in multiplex versus simplex cases is biologically meaningful, whereas that of dbSNP: rs6059655 is not, simply balancing the proportional increase of risk from dbSNP: rs9271597 (as the proportion of risk score explained by dbSNP: rs9271597 goes up, the proportion of risk score explained by other variants must go down).

Even when dbSNP: rs9271597 (as well as a correlated MHC class II variant, dbSNP: rs145954018) and the *ASIP* SNP (dbSNP: rs6059655) are removed from the CONFIRMED risk score, the score still remains significantly higher in multiplex cases than in simplex cases ($p = 0.0064$, $\hat{\beta}_{multiplex} = 0.15$ SD). These results indicate that, although the *HLA-DRB1/DQA1* locus might play a
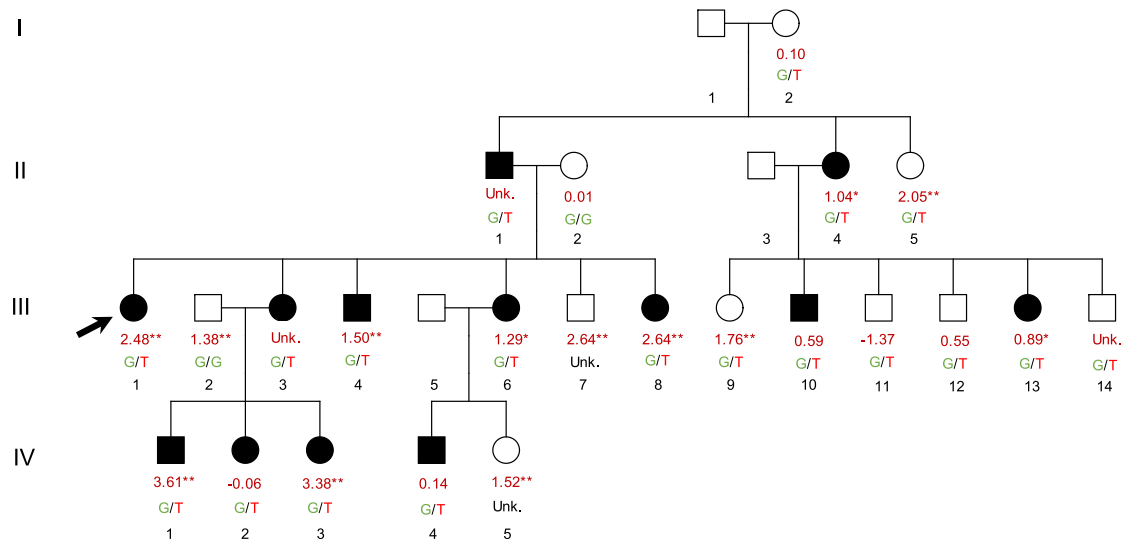
Figure 5. Annotated Pedigree for the Multiplex-Vitiligo-Affected Family with the Largest Number of Affected Relatives in Our Collection

In the largest multiplex-vitiligo-affected family in our collection, 13 individuals were affected. Symbols filled with black indicate individuals affected with vitiligo. The black arrow indicates the proband, who was genotyped genome-wide in GWAS123. Each individual is annotated with available genotype information; the top number represents the normalized FAMILY risk score (the units of the risk score are standard deviations from the mean in GWAS123 controls). * denotes individuals whose risk scores fall in the top 20th percentile and ** denotes individuals whose risk scores fall in the top 10th percentile, as defined in Table 1. The second annotation is the individual's genotype for the *FOXD3* promotor variant dbSNP: rs41285370; the rare risk allele of dbSNP: rs41285370 (T) is shown in red. Most dbSNP: rs41285370 genotypes were inferred based on linkage data.[7] Unk. = unknown, where genotype or risk score could not be inferred or calculated based on available data.

special role in vitiligo pathobiology in multiplex-affected families, the other confirmed vitiligo susceptibility loci additionally contribute to family case clustering.

Discussion

Genetic risk scores have attracted considerable attention for their potential applications as predictive or diagnostic tools, though there is some controversy surrounding their clinical utility in this regard.[20] Here, we developed and applied several different polygenic risk scores as tools to compare genetic architecture underlying two potentially different subtypes of vitiligo: simplex and multiplex. We first defined the best-performing score by comparing four different approaches for calculating a vitiligo polygenic risk score from dense genotyping data. We obtained best performance with the CONFIRMED risk score, which consists of only the most-associated GWAS variants and which outperformed risk scores that included a larger number of variants. This suggests that a large proportion of overall genetic risk for vitiligo is described by a relatively small number of variants, many captured by the lead variants at the GWAS loci and in many cases known to be pathologically causal. Thus, much of the genetic architecture of vitiligo might be defined by a limited number of variants, rather than following an infinitesimal or omnigenic model,[21] under which inclusion of a large number of additional variants would likely improve risk score performance. This further suggests that risk scores

comprised of only the most-associated variant at each locus should be considered when comparing possible risk scores for other genetically complex diseases.

By using the best-performing CONFIRMED risk score, we find that the burden of common vitiligo risk alleles is significantly higher in multiplex-affected probands than in simplex-affected subjects and that polygenic risk is over-transmitted from unaffected parents to vitiligo-affected subjects in multiplex families. Together, these findings indicate that multiplex-vitiligo-affected families segregate a high burden of the common, low-effect risk alleles identified by GWASs. Furthermore, the risk score is roughly proportional to the number of affected relatives within a family, suggesting that high polygenic risk is an important contributor to overall genetic risk for vitiligo in both smaller and larger multiplex-affected families. High polygenic burden also appears to contribute to family clustering of other traits, including bipolar disorder,[22] migraine,[23] and testicular cancer,[24] suggesting this phenomenon might be widespread among genetically complex diseases.

Although polygenic risk appears to be a major component of overall risk in multiplex-affected families, additional variants might also contribute. In our largest multiplex-affected family with 13 affected relatives, a rare, high-penetrance functional promoter variant in *FOXD3*, encoding a master regulator of melanocyte development, is carried by all affected family members.[19] Nevertheless, even within this especially large family, most affected relatives additionally have a polygenic risk score in the 80th percentile or higher, suggesting that the effect of the rare,

high-penetrance variant combines with high polygenic risk to drive overall vitiligo risk. Although all vitiligo-affected subjects in this family carry the high-penetrance *FOXD3* variant, and most also have a high polygenic risk score, three affected subjects in this family had a relatively low risk score. These three subjects demonstrate that we are unable to capture all the components of vitiligo risk, even in this special family. Furthermore, such subjects with low risk scores lend support to concerns that genetic risk scores, even those with large effects, might not be as clinically predictive at the level of the individual as previously hoped.[20]

Interestingly, we find that the MHC class II vitiligo susceptibility locus enhancer variant dbSNP: rs9271597A is significantly over-represented in multiplex-affected probands compared to simplex-affected subjects. We previously showed that dbSNP: rs9271597A is associated with increased surface expression of cell-surface HLA-DR and HLA-DQ on immune cells.[25] This suggests that HLA-related triggering of adaptive immune responses might be an especially important component of vitiligo pathobiology in multiplex-affected families. Despite apparent special importance of the MHC class II locus, the polygenic burden of the other vitiligo susceptibility loci also contributes to the phenomenon of family clustering. This has important clinical implications; whereas only about 8.6% of vitiligo cases derive from multiplex-affected families, the pathobiology of vitiligo appears to be basically the same in both familial and simplex cases. Accordingly, vitiligo therapies applicable to simplex-affected subjects are also likely to be applicable to multiplex-affected families.

We thus find that, for vitiligo, increased polygenic burden appears to contribute to the phenomenon of family clustering. Furthermore, the same set of common risk alleles that are relevant in simplex-affected subjects appear to contribute similarly to vitiligo risk in multiplex-affected families. Altogether, our findings suggest that the genetic architecture of vitiligo in multiplex-affected families is likely similar to that of sporadically-occurring vitiligo, though with additional polygenic burden. Nevertheless, not all members of multiplex-affected families have high risk scores, suggesting that some vitiligo family clusters might reflect enrichment of risk factors that are not captured by the current risk score. As such, multiplex-affected families with low risk scores seem especially likely to harbor rare, high-penetrance variants that could be identified by genetic linkage analysis. Alternatively, multiplex-vitiligo-affected families with low polygenic risk scores might reflect shared exposure to a strong environmental trigger, which could be identified by follow-up epidemiologic analysis. In either case, we suggest that families with low polygenic risk scores are atypical, and further investigation of these families might facilitate discovery of novel risk factors.

Accession Numbers

Case genotype and phenotype data for GWAS1, GWAS2, and GWAS3 subjects have been deposited in the Database of Genotypes and Phenotypes (dbGaP) under accession numbers phs000224.v1.p1, phs000224.v2.p1, phs000224.v3.p2, and phs000224.v4.p2 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000224.v3.p2). GWAS summary statistics have been deposited in the NHGRI-EBI Catalog of published genome-wide association studies (https://www.ebi.ac.uk/ega/studies/phs000224.v1.p1).

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.06.013.

Web Resources

Database of Genotypes and Phenotypes, https://www.ncbi.nlm.nih.gov
KING, http://people.virginia.edu/~wc9c/KING/manual.html
NHGRI-EBI Catalog of Published Genome-Wide Association Studies, https://www.ebi.ac.uk
PLINK-1.7, http://zzz.bwh.harvard.edu/plink/tutorial.shtml
PLINK-1.9, https://www.cog-genomics.org/plink2/
1KG Phase 3, http://www.internationalgenome.org/category/phase-3/
pROC, https://cran.r-project.org/web/packages/pROC/index.html
pTDT, https://github.com/ypaialex/ptdt
gnomAD, https://gnomad.broadinstitute.org/

References

1. Hafez, M., Sharaf, L., and Abd el-Nabi, S.M. (1983). The genetics of vitiligo. Acta Derm. Venereol. *63*, 249–251.
2. Das, S., Majumder, P.P., Majumdar, T., Haldar, B., and Rao, D.J.G.e. (1985). Studies on vitiligo. II. Familial aggregation and genetics. Genet. Epidemiol. *2*, 255–262.
3. Alkhateeb, A., Fain, P.R., Thody, A., Bennett, D.C., and Spritz, R.A. (2003). Epidemiology of vitiligo and associated autoimmune diseases in Caucasian probands and their families. Pigment Cell Res. *16*, 208–214.
4. Jin, Y., Birlea, S.A., Fain, P.R., Gowan, K., Riccardi, S.L., Holland, P.J., Mailloux, C.M., Sufit, A.J., Hutton, S.M., Amadi-Myers, A., et al. (2010). Variant of *TYR* and autoimmunity susceptibility loci in generalized vitiligo. N. Engl. J. Med. *362*, 1686–1697.
5. Jin, Y., Birlea, S.A., Fain, P.R., Ferrara, T.M., Ben, S., Riccardi, S.L., Cole, J.B., Gowan, K., Holland, P.J., Bennett, D.C., et al.

(2012). Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. Nat. Genet. *44*, 676–680.

6. Jin, Y., Andersen, G., Yorgov, D., Ferrara, T.M., Ben, S., Brownson, K.M., Holland, P.J., Birlea, S.A., Siebert, J., Hartmann, A., et al. (2016). Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. Nat. Genet. *48*, 1418–1424.

7. Alkhateeb, A., Stetler, G.L., Old, W., Talbert, J., Uhlhorn, C., Taylor, M., Fox, A., Miller, C., Dills, D.G., Ridgway, E.C., et al. (2002). Mapping of an autoimmunity susceptibility locus (*AIS1*) to chromosome 1p31.3-p32.2. Hum. Mol. Genet. *11*, 661–667.

8. Spritz, R.A., Gowan, K., Bennett, D.C., and Fain, P.R. (2004). Novel vitiligo susceptibility loci on chromosomes 7 (*AIS2*) and 8 (*AIS3*), confirmation of *SLEV1* on chromosome 17, and their roles in an autoimmune diathesis. Am. J. Hum. Genet. *74*, 188–191.

9. Majumder, P.P., Das, S.K., and Li, C.C. (1988). A genetical model for vitiligo. Am. J. Hum. Genet. *43*, 119–125.

10. Nath, S.K., Majumder, P.P., and Nordlund, J.J. (1994). Genetic epidemiology of vitiligo: Multilocus recessivity cross-validated. Am. J. Hum. Genet. *55*, 981–990.

11. Picardo, M., and Taieb, A. (2019). Vitiligo (Heidelberg, New York: Springer).

12. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.

13. Jin, Y., Roberts, G.H.L., Ferrara, T.M., Ben, S., van Geel, N., Wolkerstorfer, A., Ezzedine, K., Siebert, J., Neff, C.P., Palmer, B.E., et al. (2019). Early-onset autoimmune vitiligo associated with an enhancer variant haplotype that upregulates class II HLA expression. Nat. Commun. *10*, 391.

14. Lee, A.B., Luca, D., Klei, L., Devlin, B., and Roeder, K. (2010). Discovering genetic ancestry using spectral graph theory. Genet. Epidemiol. *34*, 51–59.

15. Hartung, J., Knapp, G., and Sinha, B.K. (2011). Statistical meta-analysis with applications (John Wiley & Sons).

16. Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J., et al.; iPSYCH-Broad Autism Group; and Psychiatric Genomics Consortium Autism Group (2017). Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat. Genet. *49*, 978–985.

17. Jostins, L., and Barrett, J.C. (2011). Genetic risk prediction in complex disease. Hum. Mol. Genet. *20* (R2), R182–R188.

18. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., and Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat. Genet. *50*, 1219–1224.

19. Alkhateeb, A., Fain, P.R., and Spritz, R.A. (2005). Candidate functional promoter variant in the *FOXD3* melanoblast developmental regulator gene in autosomal dominant vitiligo. J. Invest. Dermatol. *125*, 388–391.

20. Wald, N.J., and Old, R. (2019). The illusion of polygenic disease risk prediction. Genet. Med. *2019*, 12. https://doi.org/10.1038/s41436-018-0418-5.

21. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: From polygenic to omnigenic. Cell *169*, 1177–1186.

22. Fullerton, J.M., Koller, D.L., Edenberg, H.J., Foroud, T., Liu, H., Glowinski, A.L., McInnis, M.G., Wilcox, H.C., Frankland, A., Roberts, G., et al.; Bipolar High Risk Study Group, BiGS Consortium (2015). Assessment of first and second degree relatives of individuals with bipolar disorder shows increased genetic risk scores in both affected relatives and young At-Risk Individuals. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *168*, 617–629.

23. Gormley, P., Kurki, M.I., Hiekkala, M.E., Veerapen, K., Häppölä, P., Mitchell, A.A., Lal, D., Palta, P., Surakka, I., Kaunisto, M.A., et al.; 23andMe Research Team; and International Headache Genetics Consortium (IHGC) (2018). Common variant burden contributes to the familial aggregation of migraine in 1,589 families. Neuron *99*, 1098.

24. Loveday, C., Law, P., Litchfield, K., Levy, M., Holroyd, A., Broderick, P., Kote-Jarai, Z., Dunning, A.M., Muir, K., Peto, J., et al.; UK Testicular Cancer Collaboration, The PRACTICAL Consortium (2018). Large-scale analysis demonstrates familial testicular cancer to have polygenic aetiology. Eur. Urol. *74*, 248–252.

25. Cavalli, G., Hayashi, M., Jin, Y., Yorgov, D., Santorico, S.A., Holcomb, C., Rastrou, M., Erlich, H., Tengesdal, I.W., Dagna, L., et al. (2016). MHC class II super-enhancer increases surface expression of HLA-DR and HLA-DQ and affects cytokine production in autoimmune vitiligo. Proc. Natl. Acad. Sci. USA *113*, 1363–1368.

Supplemental Data

Family Clustering of Autoimmune Vitiligo

Results Principally from Polygenic Inheritance

of Common Risk Alleles

Genevieve H.L. Roberts, Subrata Paul, Daniel Yorgov, Stephanie A. Santorico, and Richard A. Spritz

Figure S1: Multiplex and Simplex Probands from GWAS123 are Homogeneous with Respect to European Ancestry
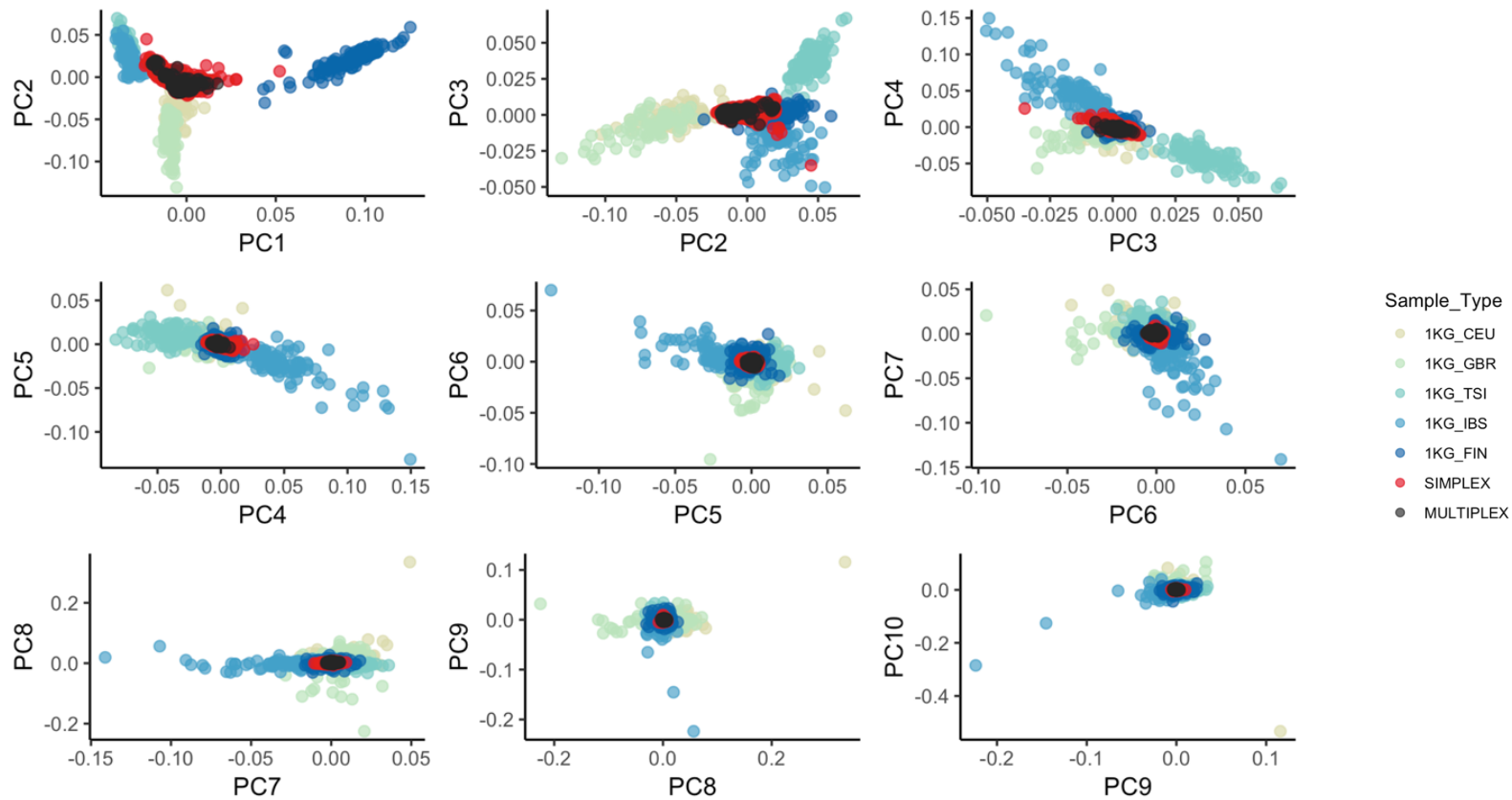
Vitiligo GWAS123 Multiplex and Simplex probands and 1000 Genomes Project European-derived (1KGP EUR) subjects plotted by ancestry-derived principal components analysis (PCA). Nine outlying 1KGP subjects were removed from plots to improve visualization of GWAS123 vitiligo cases. Each dot represents a subject and the color of each dot represents the subject's population: 1KG_CEU, 1KG Utah Resident; 1KG_GBR, 1KG British in England and Scotland; 1KG_TSI, 1KG Toscani in Italia; 1KG_IBS, 1KG Iberian Population in Spain; 1KG_FIN, 1KG Finnish in Finland; SIMPLEX, simplex cases from vitiligo GWAS123; MULTIPLEX, multiplex cases (unrelated) from vitiligo GWAS123.

Figure S2: The CONFIRMED Risk Score Distribution in GWAS123 and Independent Replication Study Subjects
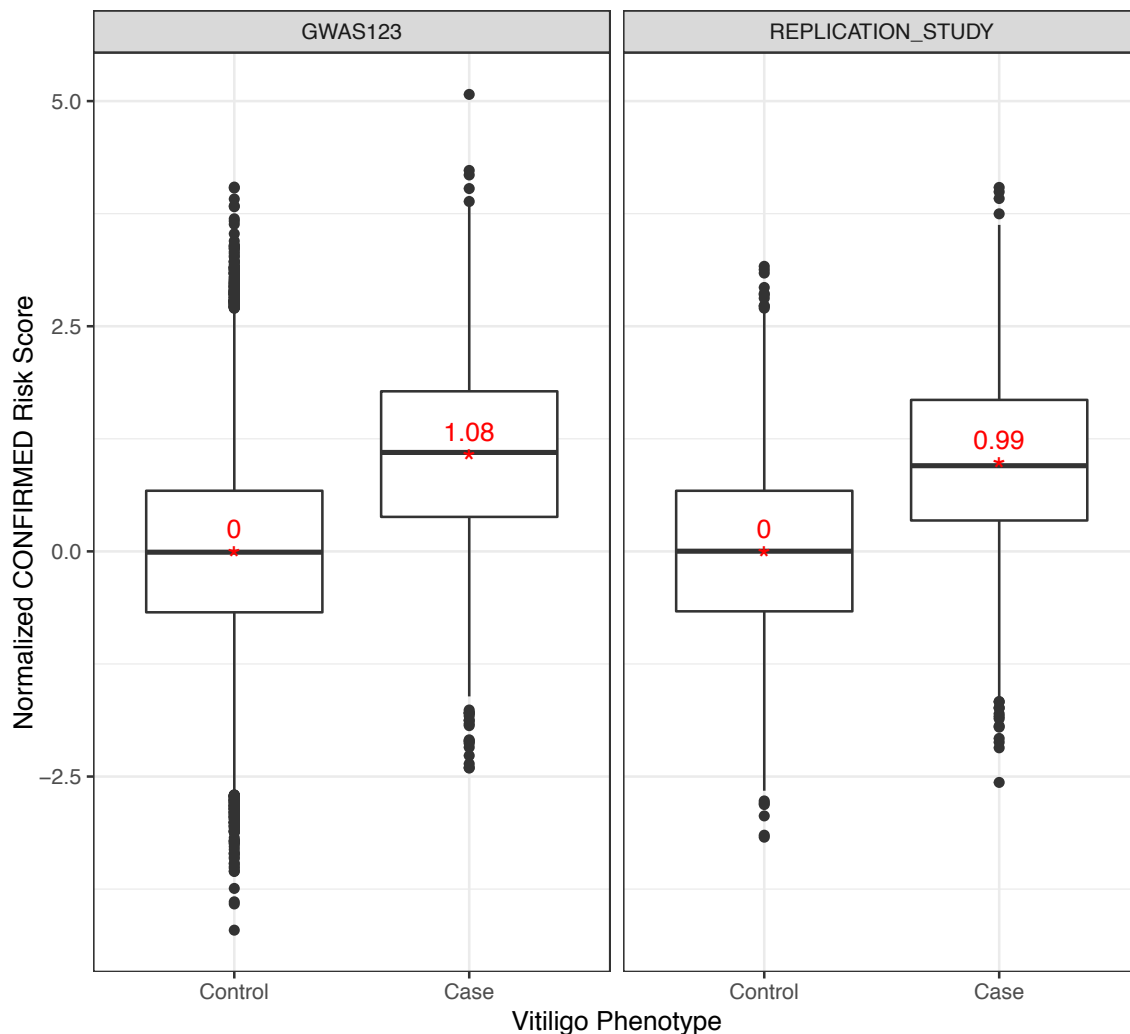
The normalized CONFIRMED risk score is plotted for vitiligo cases (n=2,841) and controls (n=37,255) in GWAS123 (left) and for cases (n=1,827) and controls (n=2,181) in the independent replication study (right). The mean and SD used for normalization were derived from GWAS123 controls; thus, the Y-axis units represent standard deviations (SD) difference from the mean risk score in GWAS123 controls. Horizontal lines denote the median, red * denotes the mean, and the value of the mean is shown in red text. Boxes denote first through third quartiles. Each vertical bar extends from the box to the largest or smallest value, no more than 1.5 times the inter-quartile range. Data beyond the vertical bars are considered outliers and are plotted individually.

Figure S3: Comparison of single-variant contributions to the CONFIRMED risk score in simplex and multiplex probands from GWAS123 + replication study
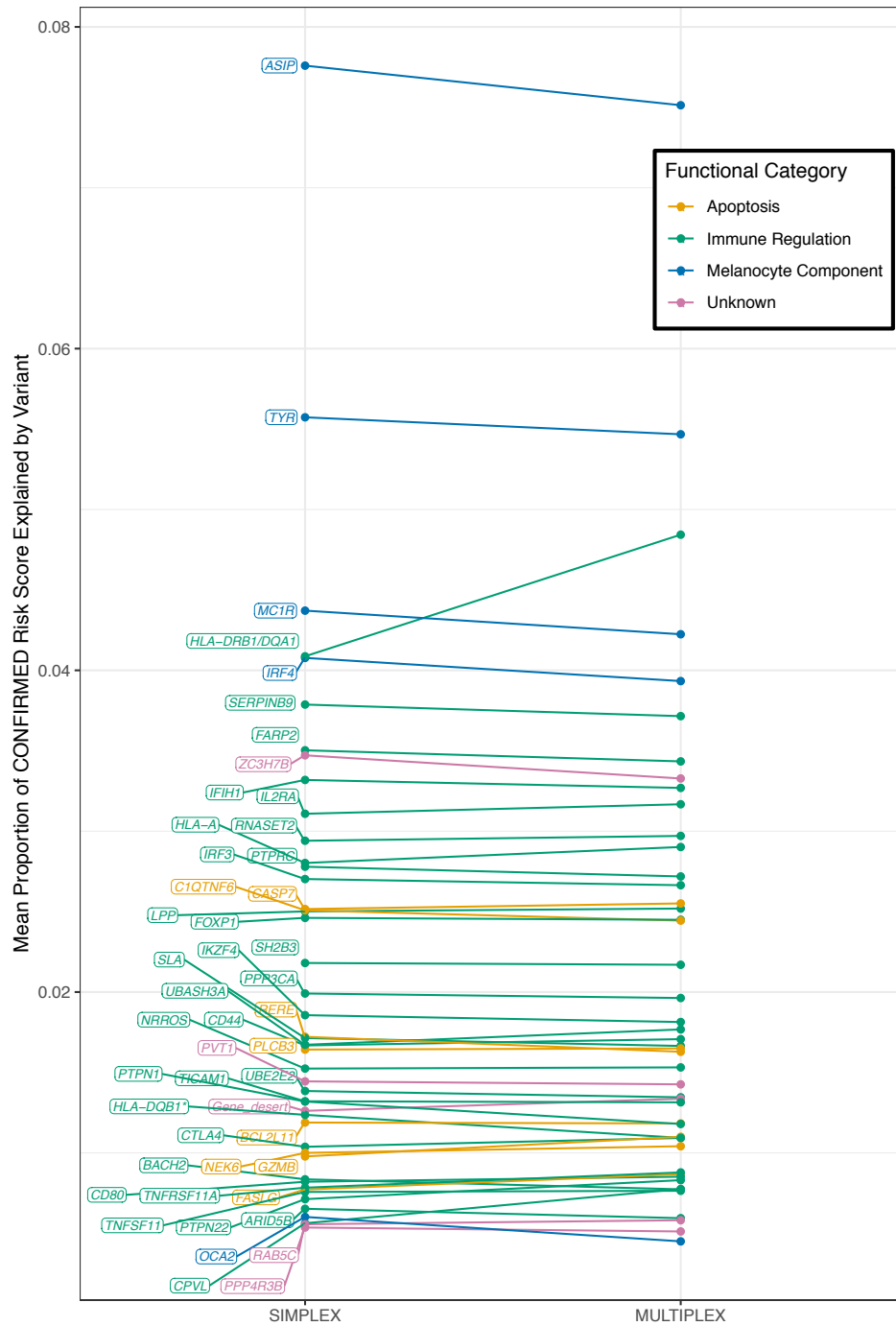
*Each point represents the mean proportion contributed to the CONFIRMED risk score by each of the 48 loci. Higher vertical position represents increased contribution to the risk score on average, with more common, high-effect-size loci generally contributing the most. Dots for the same variant are connected by a horizontal line and a larger slope represents a larger difference between mean contribution of the variant in simplex probands versus in multiplex probands. Dot and label colors represent the likely functional category of the locus, as in Figure 1. * denotes a HLA-DQB1 locus specifically associated in early-onset vitiligo cases[2]; the effect size used for the HLA-DQB1 variant here was derived from all vitiligo cases, regardless of age-of-onset. Association results for each locus are presented in Table S3.*

Table S1: Variants Used in Constructing 48-Variant CONFIRMED Risk Score

Chr.	Index variant	Position (hg19)	Locus	RA/OA	GWAS123[b] OR	Risk Score Weight[c]
1	rs301807	8484823	*RERE*	A/G	1.22	0.199
1	rs2476601	114377568	*PTPN22*	A/G	1.39	0.329
1	rs78037977	172715702	*FASLG*	G/A	1.33	0.285
1	rs16843742	198672299	*PTPRC*	T/C	1.22	0.198
2	rs10200159	55845109	*PPP4R3B*	C/T	1.48	0.392
2	rs4308124	112010486	*BCL2L11-MIR4435-2HG*	C/T	1.17	0.157
2	rs2111485	163110536	*IFIH1*	G/A	1.33	0.288
2	rs231725	204740675	*CTLA4*	A/G	1.18	0.166
2	rs41342147	242407588	*FARP2-STK25*	G/A	1.25	0.223
3	rs34080387[a]	23505187	*UBE2E2*	C/G	1.15	0.139
3	rs34346645	71557945	*FOXP1*	C/A	1.25	0.223
3	rs59374417[a]	119288414	*CD80-ADPRH*	C/A	1.37	0.315
3	rs13076312	188089254	*LPP*	T/C	1.32	0.278
3	rs6583331	196347253	*FBXO45-NRROS*	T/A	1.16	0.151
4	rs1031034	102223386	*PPP3CA*	C/A	1.16	0.151
6	rs12203592	396321	*IRF4*	C/T	1.30	0.261
6	rs78521699	2908591	*SERPINB9*	A/G	1.27	0.236
6	rs60131261	29937335	*HLA-A*	D/I	1.53	0.425
6	rs145954018[d,e]	32440321	*HLA-DQB1* & Early Vitiligo Onset[d]	D/I	2.42	0.884
6	rs9271597[e]	32591291	*HLA-DRB1/DQA1*[e]	A/T	1.77	0.571
6	rs72928038	90976768	*BACH2*	A/G	1.28	0.247
6	rs4710154[a]	167394634	*RNASET2-FGFR1OP-CCR6*	A/T	1.27	0.236
7	rs117744081	29132279	*CPVL*	G/A	1.95	0.668
8	rs10087240[f]	129012574	*PVT1*[f]	T/C	1.18	0.166
8	rs2687812	133931055	*TG-SLA-WISP1*	A/T	1.21	0.191
9	rs10986311	127071493	*NEK6*	C/T	1.16	0.148
10	rs706779	6098824	*IL2RA*	T/C	1.35	0.301
10	rs71508903	63779871	*ARID5B*	T/C	1.18	0.166

Chr.	SNP	Position	Gene	RA/OA	OR	β
10	rs12771452	115488331	*CASP7*	G/A	1.20	0.186
11	rs1043101	35274829	*CD44-SLC1A2*	G/A	1.24	0.215
11	rs12421615	64021605	*PPP1R14B-PLCB3-BAD-GPR137-KCNK4-TEX40-ESRRA-TRMT112-PRDX5*	G/A	1.15	0.139
11	rs1126809	89017961	*TYR*	G/A	1.49	0.400
11	rs11021232	95320808	Gene desert	C/T	1.38	0.322
12	rs772921[a]	56403577	*IKZF4*	T/C	1.31	0.270
12	rs10774625[a]	111910219	*SH2B3-ATXN2*	A/G	1.25	0.223
13	rs35860234	43070206	*TNFSF11*	G/T	1.16	0.148
14	rs8192917	25102160	*GZMB*	C/T	1.23	0.207
15	rs1635168	28535266	*OCA2-HERC2*	A/C	1.43	0.358
16	rs4268748	90026512	*MC1R*	T/C	1.37	0.315
17	rs11079035	40289012	*KAT2A-HSPB9-RAB5C*	A/G	1.18	0.166
18	rs8083511	60028655	*TNFRSF11A*	C/A	1.24	0.215
19	rs4807000	4831878	*TICAM1*	A/G	1.19	0.174
19	rs2304206	50168871	*SCAF1-IRF3-BCL2L12*	G/A	1.22	0.198
20	rs6059655	32665748	*RALY-EIF252-ASIP-AHCY-ITCH*	G/A	1.59	0.462
20	rs6012953	49123043	*PTPN1*	G/A	1.16	0.148
21	rs12482904	43851828	*UBASH3A*	A/T	1.43	0.358
22	rs229527	37581485	*C1QTNF6*	A/C	1.34	0.293
22	rs9611565	41767486	*ZC3H7B-TEF*	T/C	1.28	0.248

Chr., Chromosome; RA, Risk Allele; OA, Other Allele; OR, Odds Ratio.

[a]Tag SNP used for risk score calculation in place of index variant reported in GWAS123: rs34080387 in place of rs35161626 (D'= 0.997, r^2=0.993); rs59374417 in place of rs148136154 (D'=0.999 r^2=0.984); rs4710154 in place of rs2247314 (D'=0.995 r^2=0.985); rs772921 in place of rs2017445 (D'=1 r^2=0.998); rs10774625 in place of rs10774624 (D'=0.975 r^2=0.919).

[b]Published GWAS123 OR[1]

[c]Risk score weight ($\hat{\beta}$) is the ln(GWAS123 OR)

[d]Association between rs145954018 and early-onset vitiligo[2]; OR reported is from GWAS123 meta-analysis in all cases and controls without accounting for vitiligo age-of-onset.

[e]Special coding is used in risk score calculation to account for LD between rs145954018 and rs9271597, see Methods for details.

[f]Association with rs10087240 published[3] separately from GWAS123

Table S2: Variants Used in Constructing 14-Variant FAMILY Risk Score

Chr.	Locus	Position (hg19)	FAMILY Risk Score Variant	GWAS123 Top Variant (If Different from FAMILY Variant)	r^2	D'	RA/OA	Risk Score Weight[a]
1	*RERE*	8501786	rs301819	rs301807	0.97	1	A/G	0.199
1	*PTPN22*	114377568	rs2476601				A/G	0.329
2	*CTLA4*	204763882	rs3096851	rs231725	0.88	0.96	C/A	0.166
3	*FOXP1*	71573135	rs17008723	rs34346645	0.16	0.99	G/T	0.223
3	*LPP*	188089254	rs13076312				T/C	0.278
6	*HLA-A*	29942639	rs35066870	rs60131261	0.97	1.00	G/A	0.425
6	*HLA-DRB1/DQA1*	32578052	rs532098	rs9271597	0.81	0.95	A/G	0.571
10	*IL2RA*	6098824	rs706779				T/C	0.301
11	*CD44-SLC1A2*	35289819	rs1570214	rs1043101	0.99	0.99	G/A	0.215
11	*TYR*	89011046	rs1393350	rs1126809	0.93	1.00	G/A	0.400
11	Gene Desert	95311422	rs4409785	rs11021232	0.92	0.99	C/T	0.322
14	*GZMB*	25102160	rs8192917				C/T	0.207
21	*UBASH3A*	43848521	rs2839511	rs12482904	0.96	0.99	A/G	0.358
22	*C1QTNF6*	37581485	rs229527				A/C	0.293

Chr., Chromosome; r^2, SNP correlation between FAMILY and GWAS123 Top SNP; D', linkage disequilibrium between FAMILY and GWAS123 Top SNP; RA, Risk Allele; OA, Other Allele.

[a]Risk Score Weight ($\hat{\beta}$) is the ln(GWAS123 OR) of the most-associated GWAS123[1] (the same weight used in Table S1).

Table S3: CONFIRMED Risk Score Explained by Individual Variants in Multiplex and Simplex Vitiligo Cases

Variant	Locus	Mean Proportion CONFIRMED Explained in Simplex Cases	Mean Proportion CONFIRMED Explained in Multiplex Cases	$\hat{\beta}_{multiplex}$	$SE(\hat{\beta}_{multiplex})$	p value	FDR-Adjusted p value
rs301807	RERE	1.72E-02	1.63E-02	-9.40E-04	6.60E-04	0.15	0.54
rs2476601	PTPN22	7.12E-03	8.31E-03	1.20E-03	7.10E-04	0.09	0.41
rs78037977	FASLG	7.72E-03	8.68E-03	9.40E-04	6.80E-04	0.17	0.54
rs16843742	PTPRC	2.78E-02	2.72E-02	-6.00E-04	5.30E-04	0.26	0.73
rs10200159	PPP4R3B	5.36E-03	5.12E-03	-2.30E-04	6.90E-04	0.74	0.89
rs4308124	BCL2L11-MIR4435-2HG	1.19E-02	1.18E-02	-5.40E-05	5.20E-04	0.92	0.92
rs2111485	IFIH1	3.32E-02	3.27E-02	-5.10E-04	8.90E-04	0.57	0.78
rs231725	CTLA4	1.04E-02	1.09E-02	5.60E-04	5.20E-04	0.28	0.75
rs41342147	FARP2-STK25	3.50E-02	3.43E-02	-6.80E-04	4.80E-04	0.16	0.54
rs34080387	UBE2E2	1.38E-02	1.35E-02	-3.90E-04	4.60E-04	0.39	0.75
rs34346645	FOXP1	2.46E-02	2.45E-02	-1.20E-04	7.10E-04	0.87	0.92
rs59374417	CD80-ADPRH	8.21E-03	8.52E-03	2.90E-04	7.30E-04	0.69	0.87
rs13076312	LPP	2.50E-02	2.52E-02	1.70E-04	8.90E-04	0.85	0.92
rs6583331	FBXO45-NRROS	1.52E-02	1.53E-02	7.60E-05	4.90E-04	0.88	0.92
rs1031034	PPP3CA	1.99E-02	1.96E-02	-2.80E-04	4.40E-04	0.52	0.78
rs12203592	IRF4	4.08E-02	3.93E-02	-1.40E-03	5.70E-04	0.01	0.13
rs78521699	SERPINB9	3.79E-02	3.72E-02	-7.20E-04	4.80E-04	0.13	0.52
rs60131261	HLA-A	2.80E-02	2.90E-02	1.00E-03	1.30E-03	0.46	0.78
rs145954018	HLA-DQB1 & Early Vitiligo Onset	1.24E-02	1.09E-02	-1.40E-03	1.50E-03	0.34	0.75
rs9271597	HLA-DRB1/DQA1	4.09E-02	4.84E-02	7.60E-03	1.90E-03	9.6E-05	4.6E-03
rs72928038	BACH2	8.36E-03	7.74E-03	-6.30E-04	6.30E-04	0.32	0.75
rs4710154	RNASET2-FGFR1OP-CCR6	2.94E-02	2.97E-02	3.10E-04	7.10E-04	0.66	0.86
rs117744081	CPVL	5.63E-03	7.75E-03	2.10E-03	9.10E-04	0.02	0.18
rs10087240	PVT1	1.44E-02	1.43E-02	-1.90E-04	5.40E-04	0.72	0.89
rs2687812	TG-SLA-WISP1	1.71E-02	1.66E-02	-5.00E-04	6.30E-04	0.43	0.78

rs10986311	NEK6	1.00E-02	1.04E-02	4.20E-04	4.70E-04	0.37	0.75
rs706779	IL2RA	3.11E-02	3.17E-02	5.90E-04	9.50E-04	0.54	0.78
rs71508903	ARID5B	6.53E-03	5.94E-03	-5.90E-04	4.50E-04	0.19	0.57
rs12771452	CASP7	2.52E-02	2.55E-02	3.50E-04	5.20E-04	0.50	0.78
rs1043101	CD44-SLC1A2	1.67E-02	1.71E-02	3.90E-04	6.90E-04	0.57	0.78
rs12421615	PPP1R14B-PLCB3-BAD-GPR137-KCNK4-TEX40-ESRRA-TRMT112-PRDX5	1.64E-02	1.65E-02	7.20E-05	4.30E-04	0.87	0.92
rs1126809	TYR	5.57E-02	5.47E-02	-1.10E-03	1.10E-03	0.31	0.75
rs11021232	Gene desert	1.26E-02	1.34E-02	7.50E-04	8.80E-04	0.39	0.75
rs772921	IKZF4	1.86E-02	1.81E-02	-4.00E-04	8.60E-04	0.64	0.85
rs10774625	SH2B3-ATXN2	2.18E-02	2.17E-02	-1.30E-04	7.20E-04	0.86	0.92
rs35860234	TNFSF11	7.57E-03	7.64E-03	6.30E-05	4.30E-04	0.88	0.92
rs8192917	GZMB	9.78E-03	1.10E-02	1.20E-03	6.00E-04	0.04	0.26
rs1635168	OCA2-HERC2	6.02E-03	4.50E-03	-1.50E-03	6.90E-04	0.03	0.20
rs4268748	MC1R	4.37E-02	4.22E-02	-1.50E-03	8.40E-04	0.08	0.41
rs11079035	KAT2A-HSPB9-RAB5C	5.56E-03	5.81E-03	2.50E-04	4.30E-04	0.56	0.78
rs8083511	TNFRSF11A	7.84E-03	8.79E-03	9.50E-04	5.70E-04	0.09	0.41
rs4807000	TICAM1	1.32E-02	1.31E-02	-6.70E-05	5.70E-04	0.91	0.92
rs2304206	SCAF1-IRF3-BCL2L12	2.70E-02	2.66E-02	-3.80E-04	5.50E-04	0.49	0.78
rs6059655	RALY-EIF252-ASIP-AHCY-ITCH	7.76E-02	7.51E-02	-2.50E-03	7.40E-04	8.2E-04	0.02
rs6012953	PTPN1	1.32E-02	1.18E-02	-1.40E-03	4.90E-04	4.2E-03	0.07
rs12482904	UBASH3A	1.67E-02	1.77E-02	9.50E-04	1.00E-03	0.36	0.75
rs229527	C1QTNF6	2.51E-02	2.44E-02	-6.30E-04	9.60E-04	0.51	0.78
rs9611565	ZC3H7B-TEF	3.47E-02	3.33E-02	-1.40E-03	6.80E-04	0.04	0.24

SE, Standard Error

SUPPLEMENTAL REFERENCES

1. Jin, Y., Andersen, G., Yorgov, D., Ferrara, T.M., Ben, S., Brownson, K.M., Holland, P.J., Birlea, S.A., Siebert, J., Hartmann, A., et al. (2016). Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. Nat. Genet. *48*, 1418-1424.
2. Jin, Y., Roberts, G.H.L., Ferrara, T.M., Ben, S., van Geel, N., Wolkerstorfer, A., Ezzedine, K., Siebert, J., Neff, C.P., Palmer, B.E., et al. (2019). Early-onset autoimmune vitiligo associated with an enhancer variant haplotype that upregulates class II HLA expression. Nat. Commun. *10*, 391.
3. Ben, S., Jin, Y., Santorico, S.A., and Spritz, R.A. (2018). Genome-Wide Association of *PVT1* with Vitiligo. J. Invest. Dermatol. *138*, 1884-1886.