Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants

Ying Jin1,2, Genevieve Andersen1, Daniel Yorgov3, Tracey M Ferrara1, Songtao Ben1, Kelly M Brownson1, Paulene J Holland1, Stanca A Birlea1,4, Janet Siebert5, Anke Hartmann6, Anne Lienert6, Nanja van Geel7, Jo Lambert7, Rosalie M Luiten8, Albert Wolkerstorfer8, J P Wietze van der Veen8,9, Dorothy C Bennett10, Alain Taïeb11, Khaled Ezzedine11, E Helen Kemp12, David J Gawrakodger12, Anthony P Weetman12, Sulev Kõks13, Andreas Overbeck17, Alain Taïeb11, Khaled Ezzedine11, E Helen Kemp12, David J Gawrakodger12, Anthony P Weetman12, Sulev Kõks13, Andreas Overbeck17, Silvia Moretti18, Roberta Colucci18, Mauro Picardo19, Nanette B Silverberg20,21, Mats Olsson22, Yan Valle23, Igor Korobko23,24, Markus Böhm25, Henry W Lim26, Iltefat Hamzavi26, Li Zhou26, Qing-Sheng Mi26, Pamela R Fain1,2, Stephanie A Santorico1,3,27 & Richard A Spritz1,2

Vitiligo is an autoimmune disease in which depigmented skin results from the destruction of melanocytes1, with epidemiological association with other autoimmune diseases2. In previous linkage and genome-wide association studies (GWAS1 and GWAS2), we identified 27 vitiligo susceptibility loci in patients of European ancestry. We carried out a third GWAS (GWAS3) in European-ancestry subjects, with augmented GWAS1 and GWAS2 controls, genome-wide imputation, and meta-analysis of all three GWAS, followed by an independent replication. The combined analyses, with 4,680 cases and 39,586 controls, identified 23 new significantly associated loci and 7 suggestive loci. Most encode immune and apoptotic regulators, with some also associated with other autoimmune diseases, as well as several melanocyte regulators. Bioinformatic analyses indicate a predominance of causal regulatory variation, some of which corresponds to expression quantitative trait loci (eQTLs) at these loci. Together, the identified genes provide a framework for the genetic architecture and pathobiology of vitiligo, highlight relationships with other autoimmune diseases and melanoma, and offer potential targets for treatment.

In previous genome-wide linkage and association studies, we identified 27 vitiligo susceptibility loci3–6 in European-ancestry subjects, principally encoding immunoregulatory proteins, many of which are associated with other autoimmune diseases7. Several other vitiligo-associated genes encode melanocyte components that regulate normal pigmentary variation8 and in some cases are major vitiligo autoimmune antigens, with an inverse association of variation at these loci with vitiligo versus malignant melanoma4,6. To detect additional vitiligo associations with lower odds ratios, as well as uncommon risk alleles with higher odds ratios (ORs), we conducted a third GWAS (GWAS3) of European-ancestry subjects. We augmented the number of population controls in our previous GWAS1 and GWAS2 studies and performed genome-wide imputation of all three European-ancestry vitiligo GWAS. After quality control procedures, the augmented studies included 1,381 cases and 14,518 controls (GWAS1), 413 cases and 5,209 controls (GWAS2), and 1,059 cases and 17,678 controls (GWAS3), and the corresponding genomic inflation factors were 1.068, 1.059, and 1.013, respectively. We performed a fixed-effects meta-analysis of the three GWAS data sets for 8,966,411 markers (GWAS123; Online Methods). Replication used an additional 1,827 European-ancestry vitiligo cases and 2,181 controls.

Results for the three individual GWAS, the meta-analysis, and the replication study are presented in Fig. 1, Table 1, and Supplementary Table 1. Twenty-three new loci achieved genome-wide significance ($P < 5 \times 10^{-8}$) for association with vitiligo and demonstrated subsequent replication; of these loci, 21 are completely new (FASLG, PTPRC, PPAR3B, BCL2L11, FARP2–STK25, UBE2E2, FBXO45–NRROS, PPP3CA, IRF4, SERPINB9, CPVL, NEK6, ARID5B, a multigenic segment that includes BAD, TNFSF11, KAT2A–HSPA9–RAB5C, TNFRSF11A, SCAF1–IRF3–BCL2L12, a multigenic segment that includes ASIP, PTPN1, and IL1RAPL1), while two, CTLA4 and TICAM1, were suggestive in our previous studies. One previously significant locus, CLNK, was no longer significant (Supplementary Table 1). Another potential new locus, PVT1, exceeded genome-wide significance in the discovery meta-analysis ($P = 7.74 \times 10^{-9}$) but could not be successfully genotyped in the replication study and so remains uncertain. Two other loci, FLI1 and LOC101060498, exceeded genome-wide significance in the discovery meta-analysis ($P = 3.76 \times 10^{-8}$).
and $P = 3.60 \times 10^{-11}$, respectively) but did not demonstrate replication. Seven additional new loci achieved suggestive significance ($P < 1 \times 10^{-5}$) in the discovery meta-analysis (STAT4, PPARGC1B, C7orf72, PARP12, FADS2, CBFA2T3, and a chromosome 17 locus in the vicinity of AFMID) and gave evidence of replication, but failed to achieve genome-wide significance (Supplementary Table 1).

Together, the most significantly associated variants at the 48 loci (Table 1) identified by meta-analyses of the three GWAS account for 17.4% of vitiligo heritability ($h^2 \sim 0.75$). To assess whether additional independent variants at these loci might account for additional vitiligo heritability, we performed logistic regression conditional on the most significant SNP at each locus. Eight loci (FAR2–STK25, IFIH1, IL2RA, LPP, MC1R, SLA/TG, TYR, and UBA3) and the major histocompatibility complex (MHC) region showed evidence of additional independent associations, accounting for an additional 5.1% of vitiligo heritability, for a total of 22.5%. In general, the ORs for the 23 new confirmed loci were lower than those for loci detected previously, 1.15 to 1.27, except for CPVL (OR = 1.84), RALY–E2F252–ASIP–AHCY–ITCH (OR = 1.64), and IL1RA1P1 (OR = 1.77); for these three signals, the minor alleles are uncommon (minor allele frequency (MAF) = 0.03, 0.07, and 0.01, respectively) and thus were not detected in the previous GWAS because of power limitations.

To screen for functional relationships among proteins encoded at the 48 confirmed vitiligo-associated loci, we included all genes under the association peaks at these loci in unsupervised pathway analyses using g:PROFILER, PANTHER, and STRING. PANTHER and g:PROFILER identified an enriched network of BioGRID interactions, most significant for the Gene Ontology (GO) categories immune response, immune system process, positive regulation of response to stimulus, positive regulation of biological process, and regulation of response to stimulus. STRING identified a large potential interaction network (Fig. 2), with a predominance of proteins involved in immunoregulation, T cell receptor repertoire, apoptosis, antigen processing and presentation, and melanocyte function.

Considering the proteins encoded at the 23 new confirmed vitiligo candidate loci, at least 12 (CTLA4, TICAM1, PTPRC, PARP2, UBE2N2, NRROS, CPVL, ARID5B, PTPN1, TNFSF11, TNFRSF11A, IRF3, and perhaps also IL1RA1P1) have roles in immune regulation, and PPP3CA inactivates FOXP3 and is associated with canine lupus. Six proteins (FASLG, BCL2L11, BCL2L12, SERPINB9, NEK6, and BAD) are regulators of apoptosis, in particular involving immune cells. ASIP is a regulator of melanocyte gene expression, and IRF4 is a key transcription factor for both immune cells and melanocytes.

Several vitiligo-associated genes encode proteins that interact physically and functionally. BCL2L11 and BAD are binding partners that promote apoptosis, CD80 binds to CTLA4 to inhibit T cell activation, BCL2L12 binds to and neutralizes caspase-7, SERPINB9 binds to and specifically inhibits granzyme B, TNFSF11 binds to RANKL to regulate many aspects of immune cell function, including interaction of T cells and dendritic cells and thymic tolerization. Agouti signaling protein (ASIP) binds to the melanocortin-1 receptor (MC1R) to downregulate production of brown–black eumelanin, IRF4 cooperates with MITF to activate transcription of TYR. The vitiligo-associated HLA-A*02:01:01:01 subtype presents peptide antigens derived from several different melanocyte proteins, including tyrosinase (TYR), OCA2, and MC1R. Together, these relationships seem to highlight key pathways of vitiligo pathogenesis that are beginning to coalesce.
Table 1  Allelic associations at vitiligo susceptibility loci following GWAS meta-analysis and replication

| Chr. | Variant | Position (bp) (Build 37) | Locus | EA/OA | GWAS123 meta-analysis | GWAS3 replication study | GWAS123 and GWAS3 replication study meta-analysis |
|------|---------|-------------------------|-------|-------|-----------------------|------------------------|------------------------------------------------|
|      |         |                         |       |       | $P$ value             | Odds ratio             | $P$ value                                      |
|      |         |                         |       |       | $P$ value             | Odds ratio             | $P$ value                                      |
|      |         |                         |       |       | $P$ value             | Odds ratio             | $P$ value                                      |
|      |         |                         |       |       | $P$ value             | Odds ratio             | $P$ value                                      |
|      |         |                         |       |       | $P$ value             | Odds ratio             | $P$ value                                      |
|      |         |                         |       |       | $P$ value             | Odds ratio             | $P$ value                                      |

Chr., chromosome; CI, confidence interval; ND, not determined; EA, effect allele; OA, other allele. Bold highlights new significant vitiligo susceptibility loci. The chromosome 16 association peak spans a large number of genes, including MC1R.

*Heritability explained by all independent signals of the locus.
Vitiligo is epidemiologically associated with several other autoimmune, inflammatory, and immune-related disorders. Of the 7 suggestive loci, C7orf72 has been associated with Behçet’s disease, Sjögren’s syndrome, and lupus. These concordant associations for vitiligo and other autoimmune and inflammatory diseases add to those involving previously identified vitiligo susceptibility loci, which include RERE, PTPN22, IFIH1, CD80, LPP, BACH2, RNASET2–FGFR1OP–CCR6, TG–SLA, IL2RA, CD44, a chromosome 11q21 gene desert, IKZF4, SH2B3–ATXN2, UBAH3A, and C1QTNF6 (refs. 4,6). Nevertheless, in most cases, it remains uncertain whether apparent shared locus associations for different autoimmune diseases reflect shared or different underlying causal variants.

A majority of loci associated with complex traits involve causal variants that are regulatory in nature, often corresponding to apparent eQTLs. For TYR, GZMB, and MC1R, principal vitiligo risk derives from missense substitutions, whereas for OCA2 (ref. 6) and MHC class I (ref. 54) and class II (ref. 55) loci principal vitiligo risk is associated with causal variation in nearby transcriptional regulatory elements. To assess the fraction of vitiligo-associated loci for which causal variation is likely regulatory, we carried out conditional logistic regression analysis of all loci to define independent association signals, and for each signal we compiled all variants that could

Figure 2 Bioinformatic functional interaction network analysis of the proteins encoded by all positional candidate genes at all confirmed and suggestive vitiligo candidate loci. As a first step, unsupervised functional interaction network analysis was carried out using STRING v10.0 (ref. 11), considering each protein as a node and permitting ≤ 5 s-order interactions to maximize connectivity. Nodes that shared no edges with other nodes were then excluded from the network. Edge colors are from STRING: teal, interactions from curated databases; purple, experimentally determined interactions; green, gene neighborhood; blue, databases; red, gene fusions; dark blue, gene co-occurrence; pale green, text-mining; black, coexpression; lavender, protein homology. Note that SMEK2 is an alternative name for PPP4R3B.

and are key vitiligo autoantigens. IRF4 encodes a transcription factor for melanocytes, as well as lymphoid, myeloid, and dendritic cells, whose expression is controlled by alternative tissue-specific enhancers. ASIP and PPARGC1B encode paracrine regulators of melanocyte gene expression. All six loci have important roles in normal pigmentation variation, and for all six the specific alleles associated with vitiligo risk are also associated with protection from melanoma, and vice versa. The inverse genetic relationship of susceptibility to vitiligo versus melanoma suggests that vitiligo may represent enhanced immune surveillance against melanoma, consistent with the threefold reduction in melanoma incidence among patients with vitiligo and the prolonged survival of patients with melanoma who develop vitiligo during immunotherapy.

Vitiligo is epidemiologically associated with several other autoimmune diseases, including autoimmune thyroid disease, pernicious anemia, rheumatoid arthritis, adult-onset type 1 diabetes, Addison’s disease, and lupus. We searched the National Human Genome Research Institute (NHGRI)–European Bioinformatics Institute (EBI) GWAS catalog and PubMed using the 48 genome-wide significant and 7 suggestive vitiligo susceptibility loci for associations with other autoimmune, inflammatory, and immune-related disorders. Of the 23 new genome-wide significant loci for vitiligo, FASLG has been associated with celiac disease and Crohn’s disease; PTPRC has been associated with ulcerative colitis; BCL2L11 has been associated with primary sclerosing cholangitis; CTLA4 has been associated with alopecia areata, rheumatoid arthritis, autoimmune thyroid disease, myasthenia gravis, and type 1 diabetes autoantibody production; TNRFSF11A has been associated with myasthenia gravis; and ARID5B has been associated with systemic lupus erythematosus (Fig. 3). Of the seven suggestive loci, STAT4 has been associated with Behçet’s disease, Sjögren’s syndrome, and lupus, and C7orf72 has been associated with lupus. These concordant associations for vitiligo and other autoimmune and inflammatory diseases add to those involving previously identified vitiligo susceptibility loci, which include RERE, PTPN22, IFIH1, CD80, LPP, BACH2, RNASET2–FGFR1OP–CCR6, TG–SLA, IL2RA, CD44, a chromosome 11q21 gene desert, IKZF4, SH2B3–ATXN2, UBAH3A, and C1QTNF6 (refs. 4,6). Nevertheless, in most cases, it remains uncertain whether apparent shared locus associations for different autoimmune diseases reflect shared or different underlying causal variants.

A majority of loci associated with complex traits involve causal variants that are regulatory in nature, often corresponding to apparent eQTLs. For TYR, GZMB, and MC1R, principal vitiligo risk derives from missense substitutions, whereas for OCA2 (ref. 6) and MHC class I (ref. 54) and class II (ref. 55) loci principal vitiligo risk is associated with causal variation in nearby transcriptional regulatory elements. To assess the fraction of vitiligo-associated loci for which causal variation is likely regulatory, we carried out conditional logistic regression analysis of all loci to define independent association signals, and for each signal we compiled all variants that could
not be statistically distinguished. All the variants were then annotated against all available Encyclopedia of DNA Elements (ENCODE) data sets for immune-related and melanocyte-related cells (Supplementary Table 2). Overall, at approximately 58% of loci, the most significant variants (or statistically indistinguishable variants) were within a transcriptional regulatory element predicted by ENCODE data56,57. Only about 15% were in coding regions, with several resulting in missense substitutions. To further assess the general functional categories of apparent causal variants for vitiligo, we applied stratified LD score regression51 to the GWAS meta-analysis summary statistics. The greatest enrichment of heritability was observed for markers in regulatory functional categories, with considerably less enrichment for markers in protein-coding regions (Fig. 4).

We used two approaches to assess the correspondence of vitiligo association signals with expression of genes in the vicinity. We used PrediXcan51 to predict the expression of 11,553 genes in whole blood for each study subject and then tested association of predicted expression for each gene with vitiligo status. We used a Bayesian method to assess the colocalization of cis-eQTL signals in purified blood monocytes with the confirmed vitiligo association signals. The PrediXcan analysis found 83 genes for which significant differential expression was predicted in vitiligo cases versus controls after Bonferroni correction (Supplementary Table 3); of these, 75 were located within 1 Mb of one of the 48 confirmed vitiligo susceptibility loci, demonstrating highly significant enrichment (P < 0.00001) in comparison with the locations of genes that were not significant in PrediXcan analysis. The eQTL analysis found that eight of the confirmed vitiligo association signals showed significant (PP3 + PP4 > 0.99 and PP4/PP3 > 5; Online Methods) colocalization with eQTL association signals identified in purified monocytes (Supplementary Fig. 1 and Supplementary Table 4). Of the confirmed vitiligo-associated genes that could be tested using both methods, six were significant in both analyses (CASP7, HERC2–OCA2, ZC3H7B–TEF, TICAM1, RERE, and RNASET2–FGFR1OP–CCR6). For all of these genes except CASP7, one or more of the most associated SNPs not distinguishable by logistic regression were located within or very close to an ENCODE element likely to regulate gene expression in immune cell types, melanocytes, or both (Supplementary Table 2). Like a jigsaw puzzle, the pieces of the vitiligo pathogenome are thus beginning to fit together, revealing a complex network of immunoregulatory proteins, apoptotic regulators, and melanocyte components that mediate both autoimmune targeting of melanocytes in vitiligo and susceptibility to melanoma. For vitiligo, as for other complex diseases, there is enrichment of causal variation in regions that regulate gene expression. This may bode well for identifying potential therapeutic targets, as pharmacological modulation of dysregulated biological pathways may prove more tractable than attempts to target proteins affected by amino acid substitutions.

**URLs.** 1000 Genomes Project, [http://www.1000genomes.org/](http://www.1000genomes.org/); 1000 Genomes Project data, [http://www.sph.umich.edu/csg/abecasis/MACH/download/1000G-2010-08.html](http://www.sph.umich.edu/csg/abecasis/MACH/download/1000G-2010-08.html); NHGRI–EBI GWAS catalog, [http://www.ebi.ac.uk/gwas/](http://www.ebi.ac.uk/gwas/); Online Mendelian Inheritance in Man (OMIM), [http://www.ncbi.nlm.nih.gov/omim](http://www.ncbi.nlm.nih.gov/omim); PLINK, [http://pngu.mgh.harvard.edu/~purcell/plink/](http://pngu.mgh.harvard.edu/~purcell/plink/); STATA, [http://www.stata.com/](http://www.stata.com/); STRING database, [http://string-db.org/](http://string-db.org/); coloc, [https://cran.r-project.org/web/packages/coloco/index.html](https://cran.r-project.org/web/packages/coloco/index.html); GemTools, [http://wpicr.wpic.pitt.edu/WPICCompge n/GemTools/GemTools.htm](http://wpicr.wpic.pitt.edu/WPICCompgen/GemTools/GemTools.htm); IMPUTE2, [https://mathgen.stats.ox.ac.uk/impute/impute_v2.html](https://mathgen.stats.ox.ac.uk/impute/impute_v2.html); International HapMap Project, [http://hapmap.ncbi.nlm.nih.gov/](http://hapmap.ncbi.nlm.nih.gov/); PrediXcan, [https://github.com/hiroidan/PrediXcan](https://github.com/hiroidan/PrediXcan); SHAPEAT, [http://www.shapeit.fr/](http://www.shapeit.fr/); REMOAT, [http://remoat.sysbiol.cam.ac.uk/script.php](http://remoat.sysbiol.cam.ac.uk/script.php).
Accession codes. Genotype and phenotype data for GWAS1, GWAS2, and GWAS3 have been deposited in the database of Genotypes and Phenotypes (dbGaP) under phs000224.v1.p1, phs000224.v2.p1, and phs000224.v3.p1, respectively. Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
We thank the thousands of patients with vitiligo and normal control individuals around the world who participated in this study. We thank the Center for Inherited Disease Research (CIDR) for genotyping. This work used the Janus supercomputer, which is supported by the National Science Foundation (award CNS-0821794), the University of Colorado Boulder, the University of Colorado Denver, and the National Center for Atmospheric Research. The Janus supercomputer is operated by the University of Colorado Boulder. This work was supported by grants R01AR045584, R01AR056292, X01HG007484, and P30AR057212 from the US National Institutes of Health and by institutional research funding UT20-46 from the Estonian Ministry of Education and Research.

AUTHOR CONTRIBUTIONS
Y.L., G.A., and D.Y. performed statistical analyses. I.S. managed computer databases, software, and genotype data. T.M.F., S.B., G.A., and K.M.B. managed DNA samples and contributed to experimental procedures. P.H.J. managed subject coordination. S.A.B., A.H., A.L., R.M.L., A.W., J.P.W.v.d.V., N.v.G., J.L., D.C.B., A.T., K.E., E.H.K., and E.C.L. provided samples and phenotype information. S.A.S., P.R.F., and R.A.S. conceived, oversaw, and managed all aspects of the study. R.A.S. wrote the first draft of the manuscript. All authors contributed to the final manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. Genotype and phenotype data for GWAS1, GWAS2, and GWAS3 have been deposited in the database of Genotypes and Phenotypes (dbGaP) under phs000224.v1.p1, phs000224.v2.p1, and phs000224.v3.p1, respectively.
51. Finucane, H.K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).

52. Nicolae, D.L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 6, e1000888 (2010).

53. Ferrara, T.M., Jin, Y., Gowan, K., Fain, P.R. & Spritz, R.A. Risk of generalized vitiligo is associated with the common 55R-94A-247H variant haplotype of GZMB (encoding granzyme B). J. Invest. Dermatol. 133, 1677–1679 (2013).

54. Hayashi, M. et al. Autoimmune vitiligo is associated with gain-of-function by a transcriptional regulator that elevates expression of HLA-A*02:01 in vivo. Proc. Natl. Acad. Sci. USA 113, 1357–1362 (2016).

55. Cavalli, G. et al. 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 (2016).

56. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. 15, 272–286 (2014).

57. Kellis, M. et al. Defining functional DNA elements in the human genome. Proc. Natl. Acad. Sci. USA 111, 6131–6138 (2014).

58. Gamazon, E.R. et al. A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. 47, 1091–1098 (2015).
ONLINE METHODS

Subjects. The genome-wide portion of this study included unrelated cases from our three generalization vitiligo GWAS: GWAS1 (ref. 4) (n = 1,514), GWAS2 (ref. 6) (n = 450), and the current GWAS3 (n = 1,090). All cases were self-described as being of non-Hispanic/non-Latino European-derived white ancestry from North America and Europe and met strict clinical criteria for generalized vitiligo59. All controls were European-ancestry individuals not specifically known to have any autoimmune disease or malignant melanoma, for whom genome-wide genotypes were obtained from the database of Genotypes and Phenotypes (dbGaP; phs000092.v1.p1, phs000125.v1.p1, phs000138.v2.p1, phs000142.v1.l1, phs000169.v1.p1, phs000206.v5.p3, phs000237.v1.p1, phs000346.v2.p2, and phs000439.v1.p1 for GWAS1; phs000203.v1.p1 and phs000289.v2.p1 for GWAS2; phs000196.v2.p1, phs000303.v1.p1, phs000304.v2.p1, phs000368.v1.p1, phs000381.v1.p1, phs000387.v1.p1, phs000389.v1.p1, phs000395.v1.p1, phs000408.v1.p1, phs000421.v1.p1, phs000494.v1.p1, and phs000524.v2.p1 for GWAS3). Control data sets were matched to each of the three GWAS case data sets on the basis of the platforms used for genotyping. The independent replication study included 1,827 unrelated European-ancestry vitiligo cases and 2,181 unrelated European-ancestry controls not included in any of the GWAS. All subjects provided written informed consent. This study was carried out under the jurisdiction of each local institutional review board with overall oversight of the Colorado Multiple Institutional Review Board (COMIRB).

Genome-wide genotyping. Saliva specimens were obtained using a DNA self-collection kit (Oragene, DNA Genotek), and DNA was prepared using either the Maxwell apparatus/16 LEV Blood DNA kit (Promega) or the DNA Genotek Oragene Purifier protocol. DNA concentrations were measured using either the Qubit dsDNA BR Assay kit and Qubit 2.0 Fluorometer (Invitrogen) or the Promega QuantFluor ONE dsDNA kit and GloMax-Multi+ Detection System (Promega).

Genome-wide genotyping for the GWAS3 cases was performed for 716,503 variants using Illumina Human OmniExpress BeadChips by the Center for Inherited Disease Research (CIDR). Genotype data for GWAS3 was deposited in dbGaP (phs000224.v3.p1). GWAS1 (ref. 4) and GWAS2 (ref. 6) have been described previously.

Genome-wide quality control procedures. Quality control filtering of genome-wide genotype data was carried out using PLINK60, version 1.9. For each case–control data set, DNA strand calls were reversed as needed. Cases were excluded on the basis of SNP call rates <98%. SNPs were excluded on the basis of SNP call rates <98.5%, discordance between reported and observed sex, or inadvertent subject duplication, and controls were excluded on the basis of SNP call rates <98.5%. SNPs were excluded on the basis of genotype missing rate >2% for SNPs with observed MAF ≥ 0.01, and for SNPs with MAF < 0.01 the exclusion criteria were genotype missing rate >1% and <5 minor alleles observed, or significant (P < 1 × 10−4) deviation from Hardy–Weinberg equilibrium. For X-chromosome SNPs, Hardy–Weinberg equilibrium tests were performed in both females and males. Subjects were excluded on the basis of SNP call rates <95%, or discordance between reported and observed sex. Unintended duplicate samples were identified by pairwise identity-by-descent estimations (pi-hat ≥ 0.99), in which case the individual with lower SNP call rate was excluded. The final numbers of remaining cases and controls were 1,827 and 2,181, respectively, providing at least 80% power to replicate associations at P = 0.05 with Bonferroni correction for up to 48 independent tests for OR ≥ 1.23 for MAF ≥ 0.25.

Replication study genotyping and quality control procedures. For the replication study, genotyping was attempted for 379 variants using a custom Illumina GoldenGate array by CIDR. Seventy-one SNPs were excluded on the basis of genotype missing rate >2% (which includes apparent technical failures) or significant (P < 1 × 10−4) deviation from Hardy–Weinberg equilibrium. For X-chromosome SNPs, Hardy–Weinberg equilibrium tests were performed in females. Subjects were excluded on the basis of SNP call rates <95%, or discordance between reported and observed sex. Unintended duplicate samples were identified by pairwise identity-by-descent estimations (pi-hat ≥ 0.99), in which case the individual with lower SNP call rate was excluded. The final numbers of remaining cases and controls were 1,827 and 2,181, respectively, providing at least 80% power to replicate associations at P = 0.05 with Bonferroni correction for up to 48 independent tests for OR ≥ 1.23 for MAF ≥ 0.25.

Statistical analyses. To control for the effects of population stratification, we assigned cases and controls of each GWAS to homogenous clusters using GemTools69 and performed Cochran–Mantel–Haenszel (CMH) analysis to test for association for each GWAS and the combined GWAS data, with the cluster variable defined by the case–control clusters from each GWAS. After removing variants within the extended MHC, the genomic inflation factor for GWAS1, GWAS2, and GWAS3 was 1.068, 1.059, and 1.013, respectively. For the combined GWAS1–GWAS2–GWAS3 genotype data for shared SNPs, the genomic inflation factor was 1.019.

For the replication study, after quality control procedures, we compared allele frequencies for the remaining 308 SNPs in the remaining 1,827 cases and 2,181 controls using the Cochran–Armitage trend test. Odds ratios and 95% confidence limits were calculated by logistic regression analysis. We used CMH analysis to obtain odds ratios and P values for the combined GWAS plus the replication study data, with the cluster variable defined by the case–control clusters from each GWAS and the replication study data as one cluster. To analyze X-chromosome SNPs, we assumed complete X-chromosome inactivation and similar effect size for males and females, with the effect of having an A allele in a male equal to the effect of having two A alleles in a female62. We thus coded males as homozygous for the allele carried for each variant and tested for association by CMH analysis to obtain odds ratios and P values for each GWAS, the combined GWAS, and the combined GWAS plus the replication study data and by the Cochran–Armitage trend test for the replication study data.

To test the heterogeneity of associations across the three GWAS and the replication study data, we performed the Cochran Q test. The analysis was done with PLINK, version 1.07, using the odds ratios and standard errors estimated from the CMH analysis of each GWAS and from logistic regression analysis of the replication study data. The F statistic from the Q test quantifies heterogeneity and ranges from 0 to 1000% (ref. 63), with a value of 75% or greater typically taken to indicate a high degree of heterogeneity64. To test for multiple independent signals at each locus, we performed logistic regression analysis of each locus conditional on the most significantly associated
variant, including as covariates in the model the significant principal components for each GWAS derived from GemTools to control for population stratification and used a stepwise procedure to select additional variants, one by one, until no additional variants showed conditional P-values ≤ $1.0 \times 10^{-5}$. If a tested variant and the conditional variant could not improve each other significantly ($P \geq 0.05$ when comparing the two-SNP model to a single-SNP model), then both variants were considered to represent the same signal. We calculated the variance explained by a specific variant or a set of variants from the combined GWAS as the Pseudo $R^2$ of a logistic regression model that included the specific variants tested.

Bioinformatic pathway and functional enrichment analyses. To screen for functional relationships among the vitiligo candidate genes, we carried out pathway analysis of the protein products of all positional candidate genes at all 48 confirmed loci and the 7 suggestive loci using g:PROFILER, PANTHER, pathway analysis of the protein products of all positional candidate genes at all Bioinformatic pathway and functional enrichment analyses.

We carried out a proportion of the markers in the category. Enrichment per category was calculated by the ratio of the LD score software, estimated from the EUR samples in the 1000 Genomes Project Phase 1. Enrichment per category was calculated by the ratio of the estimated proportion of heritability explained by the category over the proportion of the markers in the category.

PrediXcan and monocyte eQTL colocalization analyses. We carried out a gene-based test of association of vitiligo with ‘imputed’ expression profiles for 11,553 autosomal genes in whole blood using PrediXcan. The analysis included 2,853 cases and 37,412 controls from the combined GWAS. Association testing between expression estimates for each gene and affection status for vitiligo was performed by generalized logistic regression. P values were adjusted for the number of genes tested ($n = 11,553$). NRROS, ZCH7B, TNFRSF11A, BC2L112, RALY, ASIP, OCA2, and TYR were excluded from the PrediXcan analysis because of poor prediction of gene expression in blood cells.

We derived eQTLs in peripheral blood monocytes from 414 EUR subjects with paired genotyping and gene expression data. SHAPEIT v2 was used to prephase genotypes to produce best-guess haplotypes with imputation performed using IMPUTE2 and the 1000 Genomes Project phase 1 integrated variant set version 3 (March 2012) as the reference panel. We tested for colocalization of eQTLs and vitiligo GWAS autosomal association patterns as described. Vitiligo susceptibility loci were defined by windows of robust association plus an added 100-kb buffer on both sides. eQTL probes were selected by choosing probes that resided within these windows. Probe quality annotation was performed using ReMOAT, and all probes with an annotation of ‘bad’ were removed. After removing non-autosomal loci and duplicate probe IDs, a total of 904 probes remained. All vitiligo susceptibility loci contained at least one probe with the exception of the gene desert $Y'$ of TYR, for which the only probe that intersected the locus was excluded because of a REMOAT annotation of ‘bad’. Within each locus window, all SNPs were tested for association with all probes using linear regression. P values, MAF for each SNP and respective sample sizes were used as input to test for colocalization, simultaneously testing five mutually exclusive hypotheses by generating five corresponding posterior probabilities (PPs):

- $H_0$ (PP0): there is no association with either the GWAS or the eQTL.
- $H_1$ (PP1): there is association for the GWAS only.
- $H_2$ (PP2): there is association for the eQTL only.
- $H_3$ (PP3): there is association for both the GWAS and the eQTL, but the associated variants are different for the GWAS and the eQTL.
- $H_4$ (PP4): the associated variants are the same for both the GWAS and the eQTL (colocalization).

Posterior probabilities were calculated using the R package ‘coloc’ with default settings for prior probabilities of association. Colocalization was assessed according to Guo et al., and significant colocalization was $PP_1 + PP_2 > 0.99$ and $PP_4PP_3 > 5$, and suggestive colocalization was $PP_1 + PP_2 > 0.95$ and $PP_4PP_3 > 3$.

93; significant colocalization was $PP_1 + PP_2 > 0.99$ and $PP_4PP_3 > 5$, and suggestive colocalization was $PP_1 + PP_2 > 0.95$ and $PP_4PP_3 > 3$.

- $H_0$ (PP0): there is no association with either the GWAS or the eQTL.
- $H_1$ (PP1): there is association for the GWAS only.
- $H_2$ (PP2): there is association for the eQTL only.
- $H_3$ (PP3): there is association for both the GWAS and the eQTL, but the associated variants are different for the GWAS and the eQTL.
- $H_4$ (PP4): the associated variants are the same for both the GWAS and the eQTL (colocalization).

Posterior probabilities were calculated using the R package ‘coloc’ with default settings for prior probabilities of association. Colocalization was assessed according to Guo et al., and significant colocalization was $PP_1 + PP_2 > 0.99$ and $PP_4PP_3 > 5$, and suggestive colocalization was $PP_1 + PP_2 > 0.95$ and $PP_4PP_3 > 3$.

59. Tabet, A. & Picardo, M. The definition and assessment of vitiligo: a consensus report of the Vitiligo European Task Force. Pigment Cell Res. 20, 27–35 (2007).
60. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
61. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).
62. Chang, D. et al. Accounting for eXentricities: analysis of the X chromosome in meta-analyses. Br. Med. J. 327, 557–560 (2003).
63. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558 (2002).
64. Higgins, J.P., Thompson, S.G., Deeks, J.I. & Altman, D.G. Measuring inconsistency in meta-analyses. Br. Med. J. 327, 557–560 (2003).
65. Lee, A.B., Luca, D., Klein, L., Devlin, B. & Roeder, K. Discovering genetic ancestry using spectral graph theory. Genet. Epidemiol. 34, 51–99 (2010).
66. Fairfax, B.P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 343, 1246949 (2014).
67. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014).
68. Guo, H. et al. Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. Hum. Mol. Genet. 24, 3305–3313 (2015).
69. Arloth, J., Bader, D.M., Röh, S. & Altmann, A. Re-Annotator: annotation pipeline for microarray probe sequences. PLoS One 10, e0139515 (2015).