Common variants in *FOXP1* are associated with generalized vitiligo

Ying Jin, Stanca A. Birlea, Pamela R. Fain, Christina M. Mailloux, Sheri L. Riccardi, Katherine Gowan, Paulene J. Holland, Dorothy C. Bennett, Margaret R. Wallace, Wayne T. McCormack, E. Helen Kemp, David J. Gawkrodger, Anthony P. Weetman, Mauro Picardo, Giovanni Leone, Alain Taïeb, Thomas Jouary, Khaled Ezzedine, Nanny van Geel, Jo Lambert, Andreas Overbeck & Richard A. Spritz

SUPPLEMENTARY METHODS

SUBJECTS

The initial GWA study included 1514 generalized vitiligo patients of North American and United Kingdom non-Hispanic/Latino white origin. All patients met strict clinical diagnostic criteria for generalized vitiligo[1]; subjects with other causes of localized hypopigmentation or uncertain diagnoses were excluded.

The case-control replication study included 647 additional unrelated patients and 1056 non-vitiligo controls of North American and European white origin, principally spouses of the vitiligo patients; all of these controls had no other known relatives with vitiligo. The family-based replication study included 183 unrelated simplex parents-affected offspring trios, and 332 unrelated multiplex families, consisting of two or more affected family members and relevant unaffected relatives (1383 individuals), of North American and European non-Hispanic/Latino white origin. Patients, relatives, and controls in the two replication studies provided clinical history regarding vitiligo and other autoimmune diseases.

Written informed consent was obtained from all study subjects. This study was approved by each institutional review board and was conducted according to Declaration of Helsinki principles.

GENOTYPING AND QUALITY CONTROL

Genomic DNA was prepared from peripheral-blood specimens by standard methods or from saliva specimens with use of a DNA self-collection kit using the manufacturer's instructions (Oragene, DNA Genotek). DNA concentrations were assayed by ultraviolet A_{260} spectrophotometry (NanoDrop, Thermo Scientific). Genomewide genotyping and quality control filtering of the genomewide genotyping data has been described elsewhere[2]. Briefly, we initially determined genotypes for approximately 610,000 SNPs in 1514 CEU GV cases using the Illumina 610-Quad BeadChip, and compared SNP allele frequencies to those of 2813 "public" controls genotyped using the Illumina 1M BeadChip, obtained from the Databases of Genotypes and Phenotypes (dbGaP) (n = 2731) and the Illumina iControlDB (n = 82). After stringent quality control filtering of SNP data and removal of genetic outliers and unmatchable individuals after genetic matching using GEM[3], we performed unadjusted and eigenvector-adjusted Cochran-Armitage trend tests of the remaining 520,460 SNPs in the remaining 1392 cases and 2629 controls using PLINK[4] and EIGENSTRAT[5], correcting test statistics for the remaining genomic inflation factor of 1.048 using the genomic control method[6].

Genotypes for additional SNPs across each of the candidate association regions were imputed from the GWA genotype dataset by use of MACH[7], ver.1.0 (http://www.sph.umich.edu/csg/abecasis/MACH/download/), based on phased haplotype data in the HapMap[8] CEU samples (phase II release 24, and phase III release 2). Only imputed genotypes with $r^2 > 0.3$ were used for further analyses.

In the present replication study we successfully determined genotypes for the most significant SNP from each of the seven loci that showed suggestive association in the initial GWA analysis, defined as nominal *P* values better than 10^{-4} for multiple SNPs across a contiguous genomic region, using the Sequenom MassArray iPLEX genotyping system. *FOXP1* SNP rs17008713 could not be genotyped using the MassArray system for technical reasons; accordingly, we imputed genotypes for nearby SNP rs17008723 (imputed genotype $r^2 = 0.995$), which is in almost complete LD with rs17008713 in the GWA dataset ($r^2 = 0.99$), and in the replication study we therefore genotyped rs17008723. Family-based data were subjected to Mendelian error-checking, and incompatibilities were either resolved or the incompatible individual or entire family was excluded. Two SNPs were excluded on the basis of apparent deviation from Hardy-Weinberg equilibrium in controls.

STATISTICAL ANALYSES

Statistical analyses of the GWA study have been described in detail elsewhere[2]. In the case-control replication study, after quality control filtering, we compared allele frequencies for the remaining SNPs in 647 patients and 1056 controls using the Cochran-Armitage trend test[5]. Odds ratios and 95% confidence limits were calculated by logistic regression analysis. In the family-based replication study, we calculated the association of each SNP with vitiligo using the family-based association test (FBAT)[9], version 1.5.5. Odds ratios (ORs) and 95% confidence limits were calculated by conditional logistic regression analysis, using matched case-pseudocontrol data derived from the family data[10]. To obtain combined *P* values and ORs, we performed meta-analysis using a Cochran-Mantel-Haenszel test with cases and controls from the GWA study and the case-control replication study, and cases and pseudocontrols derived from the family-based

dataset, and with just the two replication datasets. The fraction of variance accounted for by each of the confirmed GV association signals, and by all 12 loci combined, was calculated as Pseudo R^2 by logistic regression analysis of the GWA and replication study case-control dataset using STATA, assuming a multiplicative model for each SNP[11] and a polygenic multiplicative model for all 12 loci combined[12].

To test whether the P-values of the GV-associated *CCR6* SNP rs6902119 and the Crohn's disease-associated *CCR6* SNP rs2301436 were independent of each other in the vitiligo data, we used STATA to compare the fit of the logistic regression model including both loci to a model including only the conditioning locus, assuming a multiplicative genotypic effect for the high-risk allele of each locus.

REFERENCES

1. Taïeb, 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).
2. Jin, Y., *et al.* Variant of *TYR* and autoimmunity susceptibility loci in generalized vitiligo. *New Engl. J. Med*. Epub. ahead of print 2010 Apr 21; PMID: 20410501 (2010).
3. Luca, D., *et al*. On the use of general control samples for genome-wide association studies: Genetic matching highlights causal variants. *Am. J. Hum. Genet.* **82**, 453-463 (2008).
4. Purcell, S., *et al*. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet*. **81**, 559-575 (2007).
5. Price, A.L., *et al*. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904-909 (2006).
6. Devlin, B., Roeder, K., & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor. Popul. Biol.* **60**, 155-166 (2001).

7. Li, Y., Ding, J., & Abecasis, G.R. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am. J. Hum. Genet*. **79**, S2290 (2006).
8. International HapMap Consortium, *et al*. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 449, 851-861 (2007).
9. Horvath, S., *et al*. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet. Epidemiol.* **26**, 61-69 (2004).
10. Cordell, H.J., Clayton, D.G. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am. J. Hum. Genet*. **70**, 124-141 (2002).
11. Schaid, D.J. General score tests for associations of genetic markers with disease using cases and their parents. *Genet. Epidemiol*. **13**, 423-449 (1996).
12. Clayton, D.G. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PloS Genet*. **5**, e1000540 (2009).

Supplementary Table 1 Suggestive Association of Seven Loci in the GWA Dataset

Chr	SNP	nt Position (GRCh37)	Risk Allele	Risk Allele Frequency Cases	Controls	EIGENSTRAT P value	PLINK P value	Odds Ratio (95%CI)
3p13	rs11720980	71505650	A	0.371	0.324	5.30E-05	4.03E-05	1.23(1.12-1.35)
	rs6549391	71529915	A	0.710	0.674	1.00E-03	1.20E-03	1.19 (1.07-1.31)
	rs6549392	71538696	G	0.713	0.676	1.05E-03	1.10E-03	1.19 (1.07-1.31)
	rs11720523	71545170	C	0.619	0.575	2.40E-04	1.91E-04	1.2 0(1.09-1.32)
	rs6779258	71549639	T	0.618	0.571	9.57E-05	8.12E-05	1.21 (1.10-1.33)
	rs2133627	71551870	C	0.250	0.209	8.86E-05	5.54E-05	1.26 (1.13-1.40)
	rs17008713	71566417	T	0.215	0.172	6.28E-06	3.70E-06	1.32 (1.18-1.48)
	rs9866496	71568501	C	0.282	0.236	1.36E-05	1.04E-05	1.27 (1.15-1.41)
	rs1499895	71571667	A	0.607	0.563	2.34E-04	2.04E-04	1.20(1.09-1.32)
3q13.13	rs697963	107078487	T	0.992	0.982	9.58E-04	9.08E-04	2.18 (1.38-3.46)
	rs6779094	108081277	A	0.257	0.215	3.47E-05	3.96E-05	1.26 (1.13-1.41)
	rs11707076	108098637	T	0.292	0.255	5.02E-04	4.61E-04	1.21 (1.09-1.34)
	rs10511271	108114572	C	0.246	0.212	3.30E-04	4.79E-04	1.22 (1.09-1.36)
	rs4273371	108119071	T	0.549	0.500	7.12E-05	4.95E-05	1.22 (1.11-1.33)
	rs9810676	108132759	T	0.500	0.457	3.80E-04	3.01E-04	1.19 (1.09-1.31)
	rs3996020	108179729	A	0.347	0.293	2.02E-06	1.15E-06	1.28 (1.16-1.42)
	rs4299484	108189627	T	0.344	0.289	1.58E-06	9.86E-07	1.29 (1.17-1.42)
	rs2603127	108243551	A	0.265	0.213	4.50E-07	2.69E-07	1.34 (1.20-1.49)
	rs6784002	108260695	A	0.555	0.516	1.39E-03	9.30E-04	1.17 (1.07-1.29)
	rs2060986	108264948	T	0.394	0.355	1.26E-03	8.90E-04	1.18 (1.07-1.30)
	rs7633235	108454848	T	0.237	0.191	2.46E-06	1.19E-06	1.33 (1.19-1.49)
	rs9815057	108464197	T	0.234	0.189	6.59E-06	3.83E-06	1.31 (1.17-1.47)
	rs7637614	108508966	G	0.557	0.515	5.94E-04	4.43E-04	1.19 (1.08-1.30)
	rs9879707	108580778	A	0.191	0.157	2.25E-04	1.50E-04	1.27 (1.13-1.44)
6q27	rs1358786	166054922	A	0.148	0.121	7.83E-04	1.02E-03	1.25 (1.10-1.43)
	rs11755875	166216338	C	0.808	0.775	6.41E-04	7.55E-04	1.22 (1.09-1.37)
	rs9366076	167373708	C	0.818	0.787	7.15E-04	1.52E-03	1.21 (1.08-1.36)
	rs9355610	167383075	G	0.711	0.672	2.99E-04	6.33E-04	1.19 (1.08-1.32)
	rs429083	167383972	G	0.581	0.543	8.38E-04	1.34E-03	1.17 (1.07-1.28)
	rs9366078	167399512	A	0.709	0.671	3.43E-04	7.29E-04	1.19 (1.08-1.32)

	rs933243	167403873	G	0.709	0.671	3.26E-04	7.01E-04	1.19 (1.08-1.32)
	rs400837	167411008	C	0.581	0.543	9.36E-04	1.47E-03	1.17(1.06-1.28)
	rs2301436	167437988	A	0.524	0.479	2.41E-04	2.27E-04	1.19 (1.09-1.30)
	rs12529876	167461501	A	0.464	0.419	1.91E-04	1.68E-04	1.20 (1.09-1.31)
	rs10484530	167461562	A	0.446	0.405	6.61E-04	6.16E-04	1.18 (1.08-1.29)
	rs12183084	167473685	G	0.446	0.404	5.82E-04	5.47E-04	1.18 (1.08-1.29)
	rs6921588	167494397	A	0.462	0.419	4.00E-04	3.65E-04	1.19(1.08-1.29)
	rs204295	167500562	C	0.580	0.540	8.91E-04	9.84E-04	1.17 (1.07-1.28)
	rs6902119	167505791	C	0.495	0.446	7.23E-05	5.72E-05	1.21 (1.11-1.33)
	rs6456156	167522300	C	0.541	0.502	1.59E-03	1.57E-03	1.16 (1.06-1.27)
	rs4708777	167814784	A	0.922	0.904	2.86E-03	1.71E-03	1.32 (1.11-1.56)
7p21.3	rs2192346	8176301	A	0.778	0.737	5.40E-05	6.59E-05	1.26 (1.13-1.40)
	rs887848	8179846	A	0.756	0.722	1.23E-03	1.37E-03	1.20 (1.07-1.33)
	rs2110333	8185089	T	0.755	0.719	6.77E-04	8.00E-04	1.20 (1.08-1.34)
9q22.33	rs7870439	100951838	G	0.778	0.744	9.81E-04	1.02E-03	1.21 (1.08-1.34)
	rs10818610	100969348	G	0.654	0.612	5.02E-04	4.12E-04	1.19 (1.08-1.31)
	rs7868451	100987622	A	0.695	0.650	6.80E-05	8.37E-05	1.22 (1.11-1.35)
	rs1573025	100991430	G	0.838	0.804	2.63E-04	2.40E-04	1.26 (1.12-1.42)
	rs879368	100995758	C	0.748	0.705	8.42E-05	9.23E-05	1.24 (1.11-1.37)
	rs7853442	101008885	A	0.693	0.649	8.94E-05	1.08E-04	1.22 (1.10-1.34)
	rs10115971	101024887	A	0.834	0.801	6.61E-04	6.36E-04	1.24 (1.10-1.39)
	rs10818692	101026301	A	0.834	0.801	5.48E-04	5.37E-04	1.24 (1.10-1.39)
	rs10760233	101034626	G	0.745	0.708	7.61E-04	7.43E-04	1.20 (1.08-1.32)
	rs4743196	101049252	G	0.662	0.623	6.63E-04	8.05E-04	1.18 (1.07-1.29)
12q13.2	rs11171710	56368078	G	0.582	0.541	5.10E-04	4.94E-04	1.19 (1.08-1.30)
	rs773107	56369506	G	0.376	0.319	1.03E-06	6.60E-07	1.28 (1.17-1.41)
	rs10876864	56401085	G	0.467	0.411	5.57E-06	2.74E-06	1.25 (1.14-1.37)
	rs1701704	56412487	C	0.389	0.329	2.54E-07	1.66E-07	1.30 (1.18-1.42)
	rs705708	56488913	A	0.507	0.463	1.53E-04	2.15E-04	1.20 (1.09-1.31)
	rs10783779	56491880	G	0.445	0.405	4.34E-04	6.25E-04	1.18 (1.08-1.30)
12q24.12	rs12311063	110557312	A	0.656	0.619	1.49E-03	1.73E-03	1.17 (1.06-1.29)
	rs12313068	110597304	C	0.865	0.836	9.70E-04	6.98E-04	1.26(1.11-1.44)
	rs11065287	110609714	T	0.342	0.296	8.82E-05	3.53E-05	1.24 (1.12-1.36)
	rs12318836	110642190	A	0.883	0.855	7.49E-04	4.76E-04	1.29 (1.12-1.48)

rs10774599	110697448	T	0.667	0.630	2.67E-03	1.53E-03	1.17 (1.07-1.29)
rs3026445	110723203	T	0.669	0.631	4.25E-04	1.26E-03	1.18 (1.07-1.29)
rs7957299	111002311	G	0.361	0.319	3.81E-04	1.83E-04	1.21 (1.10-1.34)
rs7975139	111221131	T	0.908	0.884	1.64E-03	1.24E-03	1.30 (1.11-1.52)
rs850511	111239821	A	0.482	0.441	9.50E-04	4.84E-04	1.19 (1.08-1.30)
rs7970490	111756438	A	0.741	0.701	4.75E-04	3.60E-04	1.21 (1.09-1.34)
rs3847953	111765464	G	0.747	0.711	1.32E-03	1.12E-03	1.19 (1.07-1.31)
rs3184504	111884608	T	0.544	0.490	1.47E-05	6.91E-06	1.24 (1.13-1.36)
rs653178	112007756	G	0.545	0.492	2.05E-05	9.58E-06	1.24 (1.13-1.35)
rs11065987	112072424	G	0.475	0.432	6.75E-04	3.72E-04	1.19 (1.08-1.30)
rs17696736	112486818	G	0.480	0.440	1.41E-03	8.22E-04	1.17 (1.07-1.28)
rs11066188	112610714	A	0.462	0.423	2.13E-03	1.23E-03	1.17 (1.07-1.28)
rs11066320	112906415	A	0.475	0.436	2.28E-03	1.42E-03	1.17 (1.06-1.27)
rs233716	113039943	G	0.437	0.397	1.85E-03	9.82E-04	1.17 (1.07-1.28)

Listed SNPs are from the 2000 with highest-ranked PLINK P values in the GWA study

Supplementary Table 2 Genotype counts for SNPs in novel candidate GV susceptibility loci

SNP					GWA Study		Replication 1		Replication 2
Chr	Locus Region	ID	Location (nt)	Genotypes	Cases	Controls	Cases	Controls	Cases
Replicated loci									
3p13	*FOXP1*	rs17008713	71566417						
				GG	61 (4.4)	83 (3.1)	25 (4.1)	25 (2.8)	30 (3.2)
				GT	476 (34.2)	735 (28.0)	210 (34.8)	253 (28.3)	313 (33.6)
				TT	855 (61.4)	1811 (68.9)	369 (61.1)	617 (68.9)	590 (63.2)
6q27	*CCR6*	rs6902119	167505791						
				CC	361 (25.9)	527 (20.0)	124 (21.3)	158 (18.9)	243 (24.9)
				CT	655 (47.1)	1290 (49.1)	353 (60.5)	459 (54.8)	495 (50.7)
				TT	376 (27.0)	812 (30.9)	106 (18.2)	220 (26.3)	239 (24.4)
Unconfirmed loci									
3q13.13	*MYH15*	rs2603127	108243551						
				AA	92 (6.6)	118 (4.5)	20 (3.5)	30 (3.6)	67 (7.6)
				AG	553 (39.7)	880 (33.5)	167 (28.8)	217 (26.1)	285 (32.2)
				GG	747 953.7)	1624 (61.9)	392 (67.7)	585 (70.3)	533 (60.2)
Excluded loci									
7p21.3	*ICA1*	rs2192346	8176301						
				AA	840 (60.4)	1413 (53.8)	347 (57.0)	513 (57.6)	581 (59.8)
				AG	485 (34.9)	1049 (39.9)	221 (36.3)	332 (37.3)	350 (36.1)
				GG	66 (4.7)	166 (6.3)	40 (6.6)	45 (5.1)	40 (4.1)
9q22.33	*TBC1D2*	rs7868451	100987622						
				AA	686 (49.3)	1127 (42.9)	214 (37.3)	370 (46.8)	418 (44.0)
				AG	563 (40.4)	1162 (44.2)	288 (50.3)	332 (42.0)	422 (44.5)
				GG	143 (10.3)	340 (12.9)	71 (12.4)	88 (11.1)	109 (11.5)

Supplementary Table 3 Replication analysis of novel candidate GV susceptibility loci – Excluded loci

		GWA Study				Replication 1		Replication 2		Meta-analysis Replication 1 + Replication 2		Meta-analysis GWA + Replication 1 + Replication 2	
SNP	Risk allele	$AF_{RA,}$ cases	$AF_{RA,}$ controls	PLINK *P*	OR	*P*	OR	*P*	OR	*P*	OR	*P*	OR
rs2192346	A	0.778	0.734	6.59×10^{-5}	1.26	0.510	0.94	0.572	0.90	0.627	0.96	1.42×10^{-3}	1.16
rs7868451	A	0.695	0.650	8.37×10^{-5}	1.22	3.45×10^{-3}	0.79	0.346	1.01	0.016	0.82	0.022	1.10

SNP rs2192346 and rs7868451 were located at nt 8,176,301in the *ICA1* region of 7p21.3, and nt 100,987,622 in the *TBC1D2* region of 9q22.33, respectively. SNP nucleotide positions are from GRCh37 and genes in close proximity to the designated SNP are denoted. AF_{RA}, allele frequency of the risk allele EIGENSTRAT GWA *P*-values for SNPs rs2192346 and rs7868451 were 5.40×10^{-5}, and 6.80×10^{-5}, respectively. PLINK and EIGENSTRAT GWA test statistics were calculated and adjusted for the genomic inflation factor 1.048 as described in the Supplementary Methods. The Bonferroni adjusted significance threshold for the combined replication stage 1 + 2 meta-analysis was $P < 1.00 \times 10^{-2}$, and the significance threshold for the overall combined GWA + replication stage 1 + 2 meta-analysis was $P < 5 \times 10^{-8}$.