ARTICLE

Sequencing-based approach identified three new susceptibility loci for psoriasis

Yujun Sheng[1,2,3,*], Xin Jin[4,5,*], Jinhua Xu[3,*], Jinping Gao[2,6,7,*], Xiaoqing Du[2,6,7], Dawei Duan[2,6,7], Bing Li[2,6,7], Jinhua Zhao[2,6,7], Wenying Zhan[2,6,7], Huayang Tang[2,6,7], Xianfa Tang[2,6,7], Yang Li[2,6,7], Hui Cheng[2,6,7], Xianbo Zuo[2,6,7], Junpu Mei[4], Fusheng Zhou[2,6,7], Bo Liang[2,6,7], Gang Chen[2,6,7], Changbing Shen[2,6,7], Hongzhou Cui[2,6,7], Xiaoguang Zhang[2,6,7], Change Zhang[2,6,7], Wenjun Wang[2,6,7], Xiaodong Zheng[2,6,7], Xing Fan[2,6,7], Zaixing Wang[2,6,7], Fengli Xiao[2,6,7], Yong Cui[2,6,7], Yingrui Li[4], Jun Wang[4,8,9], Sen Yang[2,6,7], Lei Xu[10], Liangdan Sun[2,6,7] & Xuejun Zhang[1,2,3,6,7]

In a previous large-scale exome sequencing analysis for psoriasis, we discovered seven common and low-frequency missense variants within six genes with genome-wide significance. Here we describe an in-depth analysis of noncoding variants based on sequencing data (10,727 cases and 10,582 controls) with replication in an independent cohort of Han Chinese individuals consisting of 4,480 cases and 6,521 controls to identify additional psoriasis susceptibility loci. We confirmed four known psoriasis susceptibility loci (*IL12B*, *IFIH1*, *ERAP1* and *RNF114*; $2.30 \times 10^{-20} \leq P \leq 2.41 \times 10^{-7}$) and identified three new susceptibility loci: 4q24 (*NFKB1*) at rs1020760 ($P = 2.19 \times 10^{-8}$), 12p13.3 (*CD27-LAG3*) at rs758739 ($P = 4.08 \times 10^{-8}$) and 17q12 (*IKZF3*) at rs10852936 ($P = 1.96 \times 10^{-8}$). Two suggestive loci, 3p21.31 and 17q25, are also identified with $P < 1.00 \times 10^{-6}$. The results of this study increase the number of confirmed psoriasis risk loci and provide novel insight into the pathogenesis of psoriasis.

[1] Department of Dermatology, No. 2 Hospital, Anhui Medical University, Hefei, Anhui 230601, China. [2] Institute of Dermatology and Department of Dermatology, No. 1 Hospital, Anhui Medical University, Hefei, Anhui 230022, China. [3] Department of Dermatology, Huashan Hospital of Fudan University, Shanghai 200040, China. [4] BGI-Shenzhen, Shenzhen 518083, China. [5] School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 518055, China. [6] Department of Dermatology and Venereology, Anhui Medical University, Hefei, Anhui 230032, China. [7] Key Laboratory of Dermatology, Anhui Medical University, Ministry of Education, China, Hefei, Anhui 230032, China. [8] Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen N, Denmark. [9] Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark. [10] Department of Computer Science and Engineering, Rm1028 Ho Sin-Hang Engineering Building, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to X.J.Z (email: ayzxj@vip.sina.com) or to L.S. (email: ahmusld@163.com) or to L.X. (email: lxu@cse.cuhk.edu.hk).

Psoriasis is a common, chronic, inflammatory disease of the skin, affecting approximately 2% of the population worldwide[1]. Our understanding of the genetic architecture of psoriasis has increased rapidly due to genome-wide association studies (GWASs), which have identified more than 40 psoriasis susceptibility genes/loci[2–13]. However, the combined effect of these loci does not fully account for the observed genetic susceptibility to psoriasis, indicating that additional genetic factors remain to be discovered. Furthermore, most of the associated variants emerging from GWASs lie within noncoding regions[14].

In our previous sequencing analysis, which aimed to investigate the contribution of functional coding variants to the genetic component of psoriasis by using 10,727 cases and 10,582 controls, 7 common and low-frequency missense single-nucleotide variants (SNVs) within *IL23R*, *GJB2*, *LCE3D*, *ERAP1*, *CARD14* and *ZNF816A* were identified as being associated with psoriasis[15]. However, single-variant and gene-based association analyses of non-synonymous SNVs failed to identify new genes associated with psoriasis, indicating that coding variants, at least non-synonymous ones with low and rare frequencies might make a limited contribution to the overall genetic risk of psoriasis.

In this study, we performed an in-depth data analysis focusing on noncoding SNVs, which were derived from the flanking regions (within ~200 bp) of target genes in the exome sequencing and targeted sequencing stages (10,727 cases and 10,582 controls)[15], and conducted a replication study in an independent cohort of Han Chinese individuals (4,480 cases and 6,521 controls) to explore additional common susceptibility variants for psoriasis. We confirmed four known psoriasis susceptibility loci (*IL12B*, *IFIH1*, *ERAP1* and *RNF114*) and identified three new susceptibility loci (*NFKB1*, 12p13.3 and 17q12) as well as two suggestive loci (3p21.31 and 17q25) for psoriasis. These findings reveal new genetic susceptibility factors and suggest several new biological pathways related to psoriasis.

Results

Exome sequencing and targeted sequencing. During exome sequencing of samples from 781 cases and 676 controls, we identified 518,308 SNVs. Of these variants, 133,671 were noncoding SNVs with an MAF $> 0.5\%$. Through the targeted sequencing of 1,362 genes from 9,946 cases and 9,906 controls,

14,365 high-confidence noncoding SNVs were detected[15]. By subjecting a combined data set of 1,362 genes (from 10,727 psoriasis cases and 10,582 controls) obtained through these two steps to a trend test, the SNV rs1020760, which is located within *NFKB1*, reached genome-wide significance ($P_{\text{exome-target}} = 2.19 \times 10^{-8}$, odds ratio (OR) = 1.12) and the SNV rs1609798 at this locus showed suggestive association ($P_{\text{exome-target}} = 9.87 \times 10^{-8}$, OR = 1.12) (Table 1) as determined by the logistic regression (additive model). Linkage disequilibrium (LD) analysis showed that these two significant SNVs represented a signal showing a strong LD (D' = 0.98, $r^2 = 0.66$). Further conditional analysis indicated that rs1020760 ($P_{\text{condition}} = 3.52 \times 10^{-2}$, OR = 1.08) was more significant than rs1609798. ($P_{\text{condition}} = 1.64 \times 10^{-1}$, OR = 1.05) (Supplementary Table 1). In addition, we confirmed three previously reported psoriasis susceptibility genes: *IL12B* (rs2288831, $P_{\text{exome-target}} = 2.30 \times 10^{-20}$, OR = 0.83), *IFIH1* (rs13431841, $P_{\text{exome-target}} = 2.96 \times 10^{-9}$, OR = 0.83) and *ERAP1* (rs27043, $P_{\text{exome-target}} = 6.50 \times 10^{-12}$, OR = 0.87) with logistic regression (additive model). We also observed 22 additional SNVs showing suggestive associations ($P_{\text{target}} < 0.01$) (Table 2 and Supplementary Table 2), which were selected for validation via genotyping. Genotype information of every sample on reported associated SNP locations were available in Supplementary Data 1.

Genotyping validation. We genotyped the 22 SNVs showing suggestive evidence ($P_{\text{target}} < 0.01$) in additional 4,480 psoriasis cases and 6,521 controls. After we combined the genotyping data from these three stages, we found that three additional SNVs at two loci reached genome-wide significance ($P_{\text{combined}} = 5.00 \times 10^{-8}$, Table 2) by using logistic regression (additive model): 12p13.3 (rs758739, $P_{\text{combined}} = 4.08 \times 10^{-8}$, OR = 0.91; rs2243750, $P_{\text{combined}} = 4.38 \times 10^{-8}$, OR = 0.91) and 17q12 (rs10852936, $P_{\text{combined}} = 1.96 \times 10^{-8}$, OR = 1.10). In addition, SNV rs12936231 within 17q12 also showed a strong association with psoriasis (rs12936231, $P_{\text{combined}} = 5.02 \times 10^{-8}$, OR = 1.10). LD analysis revealed that the two SNVs at 12p13.3 represented a single signal displaying a strong LD (D' = 0.98, $r^2 = 0.78$), whereas the two SNVs at 17q12 were in complete LD (D' = 1.00, $r^2 = 1.00$). In the validation stage, we also identified two additional loci exhibiting suggestive evidence of association

Table 1 | Association evidence for five SNVs obtained in the exome sequencing and targeted sequencing stages through logistic regression (additive model).

Chr	Gene of interest	Variant ID	Stage*	Allele	MAF Cases	MAF Controls	P-value	OR (95% CI)	Combined P-value	Combined OR (95% CI)
4q24	*NFKB1*	rs1020760	i	G/C	0.4351	0.4371	9.49×10^{-1}	0.99 (0.95–1.24)	2.19×10^{-8}	1.12 (1.08–1.17)
			ii	G/C	0.4634	0.4352	1.76×10^{-8}	1.12 (1.02–1.26)		
4q24	*NFKB1*	rs1609798	i	T/C	0.3675	0.3380	2.31×10^{-1}	1.14 (0.84–1.14)	9.87×10^{-8}	1.12 (1.08–1.17)
			ii	T/C	0.3698	0.3433	9.87×10^{-8}	1.12 (1.08–1.17)		
2q24	*IFIH1*	rs13431841	i	A/G	0.0625	0.1184	2.56×10^{-1}	0.50 (0.15–1.70)	2.96×10^{-9}	0.83 (0.78–0.88)
			ii	A/G	0.1021	0.1208	3.84×10^{-9}	0.83 (0.78–0.88)		
5q31.1–q33.1	*IL12B*	rs2288831	i	C/T	0.3684	0.4324	1.68×10^{-2}	0.77 (0.62–0.95)	2.30×10^{-20}	0.83 (0.80–0.86)
			ii	C/T	0.3986	0.4434	3.11×10^{-19}	0.83 (0.80–0.87)		
5q15	*ERAP1*	rs27043	i	A/G	0.4681	0.4803	5.61×10^{-1}	0.95 (0.81–1.12)	6.50×10^{-12}	0.87 (0.84–0.91)
			ii	A/G	0.4635	0.4986	4.33×10^{-12}	0.87 (0.83–0.90)		

CI, confidence interval; MAF, mutation annotation format; OR, odds ratio; SNV, single-nucleotide variant.
*Stage i: exome sequencing in 1,457 individuals; stage ii: targeted sequencing in 19,852 individuals.

Table 2 | Association results from the exome sequencing, targeted sequencing, replication and combined analyses by using logistic regression (additive model).

Chr	Gene of interest	Variant ID	Stage*	Allele	MAF		P-value	OR (95% CI)	Combined	
					Cases	Controls			P-value	OR (95% CI)
12p13.3	Multiple genes	rs758739	i	G/C	0.2662	0.2957	2.08×10^{-1}	0.86 (0.72-1.04)	4.08×10^{-8}	0.91 (0.89-0.97)
			ii	G/C	0.2681	0.2822	1.72×10^{-3}	0.93 (0.89-0.97)		
			iii	G/C	0.2642	0.2911	4.51×10^{-6}	0.87 (0.83-0.93)		
12p13.3	Multiple genes	rs2243750	i	C/T	0.3026	0.3229	2.48×10^{-1}	0.91 (0.76-1.05)	4.38×10^{-8}	0.91 (0.89-0.97)
			ii	C/T	0.3265	0.3436	4.70×10^{-4}	0.93 (0.89-0.97)		
			iii	C/T	0.3056	0.3308	3.49×10^{-5}	0.89 (0.84-0.94)		
17q12	IKZF3	rs10852936	i	T/C	0.2795	0.2563	2.42×10^{-1}	1.13 (0.88-1.25)	1.96×10^{-8}	1.10 (1.05-1.15)
			ii	T/C	0.2972	0.2778	2.31×10^{-5}	1.10 (1.05-1.15)		
			iii	T/C	0.2939	0.2722	2.26×10^{-4}	1.11 (1.05-1.18)		
17q12	IKZF3	rs12936231	i	G/C	0.3019	0.2537	4.07×10^{-1}	1.27 (1.11-1.96)	5.02×10^{-8}	1.10 (1.04-1.17)
			ii	G/C	0.3015	0.2837	9.24×10^{-5}	1.09 (1.04-1.14)		
			iii	G/C	0.2928	0.2726	6.87×10^{-4}	1.11 (1.04-1.17)		
20q13.2	RNF114	rs4647954	i	C/T	0.4234	0.4021	2.49×10^{-1}	1.09 (0.94-1.27)	2.41×10^{-7}	1.09 (1.05-1.14)
			ii	C/T	0.4316	0.4105	2.27×10^{-5}	1.09 (1.05-1.14)		
			iii	C/T	0.4267	0.4086	6.04×10^{-3}	1.08 (1.02-1.14)		
3p21.31	CDCP1	rs1863837	i	C/T	0.2412	0.2494	6.82×10^{-1}	0.96 (0.74-1.09)	3.91×10^{-7}	0.90 (0.86-0.95)
			ii	C/T	0.2240	0.2414	8.87×10^{-5}	0.91 (0.86-0.95)		
			iii	C/T	0.2144	0.2356	2.06×10^{-3}	0.89 (0.82-0.96)		
17q25	TRIM47	rs3744017	i	A/G	0.1837	0.1490	1.34×10^{-2}	1.26 (1.05-1.56)	5.83×10^{-7}	1.12 (1.05-1.17)
			ii	A/G	0.1595	0.1462	2.70×10^{-4}	1.11 (1.05-1.17)		
			iii	A/G	0.1656	0.1522	5.70×10^{-3}	1.11 (1.03-1.19)		

CI, confidence interval; MAF, mutation annotation format; OR, odds ratio.
*Stage i: exome sequencing in 1,457 individuals; stage ii: targeted sequencing in 19,852 individuals; stage iii: genotyping in 11,001 individuals.

through logistic regression (additive model, Table 2): 3p21.31 (rs1863837, $P_{combined} = 3.91 \times 10^{-7}$, OR = 0.90) and 17q25 (rs3744017, $P_{combined} = 5.83 \times 10^{-7}$, OR = 1.12) and confirmed the association of RNF114 with psoriasis (rs4647954, $P_{combined} = 2.41 \times 10^{-7}$, OR = 1.09).

Discussion

In the current study, we performed an in-depth data analysis focusing on noncoding SNVs based on large-scale sequencing data (from 10,727 cases and 10,582 controls)[15]. Furthermore, we performed a replication study in an independent cohort of Han Chinese individuals (4,480 cases and 6,521 controls). We replicated four known susceptibility loci (IL12B, IFIH1, ERAP1 and RNF114) and identified three novel psoriasis susceptibility loci: NFKB1 (rs1020760), 12p13.3 (rs758739) and 17q12 (rs10852936) (Supplementary Fig. 1). In addition, we identified two suggestive loci (3p21.31 and 17q25). Although we achieved good sequencing coverage and depth[15], no functional coding variants (non-synonymous) were identified that showed significance within these new susceptibility loci.

At 4q24, we identified two noncoding variants (rs1020760 and rs1609798) within NFKB1 (Supplementary Fig. 1a). NFKB1 (nuclear factor of kappa light polypeptide gene enhancer in B-cells 1) encodes a 105 kDa protein that functions as a Rel protein-specific transcription inhibitor and a 50 kDa protein that functions as a DNA binding subunit of the NF-kappa-B (NF-κB) protein complex[16]. NF-κB is one of the most important regulators of proinflammatory gene expression and has an important role in balancing growth and differentiation in the epidermis[17]. The synthesis of cytokines such as tumour necrosis factor α (TNF-α), interleukin-6 (IL-6), IL-1b and IL-8 is mediated by NF-κB. In psoriasis, many cytokine transcription factors and inflammatory mediators released from chronic inflammatory cells have important roles in modulating and regulating keratinocyte differentiation and proliferation[18]. As NF-κB regulates cytokine gene expression, it is likely that NF-κB is one of the key factors involved in the pathogenesis of this disease. Another gene, NFKB1A (14q13.2), has also been described to be strongly associated with psoriasis[2,9]. Furthermore, we performed a gene expression analysis based on public psoriasis databases[19] and found that NFKB1 is upregulated in skin samples from psoriatic patients compared with normal controls (both comparing skin samples from cases (PP) versus normal skin from controls (NN) ($P = 2.33 \times 10^{-24}$) and comparing affected (PP) versus unaffected (PN) skin samples obtained from cases ($P = 3.28 \times 10^{-21}$)). In addition, interrogation of the ENCODE database revealed that rs1609798 lies in a DH site in the last intron of NFKB1 in human umbilical vein endothelial cells (HUVECs), skeletal muscle myoblasts (HSMMs) and embryonic stem cells (H1s).

Candidate genes lying close to the associated SNVs (rs758739 and rs2243750) at 12p13.3 included CD27 (CD27 molecule), CHD4 (chromodomain helicase DNA binding protein 4) and LAG3 (lymphocyte activation gene 3) (Supplementary Fig. 1b). These genes were determined to be candidates based on their biological implications for psoriasis. CD27, a member of the TNF receptor superfamily, is a transmembrane receptor required for the generation and long-term maintenance of T-cell immunity[20]. It binds to the ligand CD70 and provides co-stimulatory signals

3

for T-, B- and NK-cell activation. Furthermore, it enhances T-cell survival and effector function, NK-cell function, B-cell differentiation and plasma cell function[21,22]. *CHD4* is a catalytic subunit of the NuRD (nucleosome remodelling and deacetylase) complex. Recently, it has emerged that *CHD4* has important roles in the DNA damage response, transcriptional regulation, cell cycle progression and the maintenance of genomic integrity[23]. *LAG3*, a ligand for MHC class II receptors, is closely related to *CD4* and downregulates the activated T-cells on which it is expressed through a high-affinity interaction with its receptor; this interaction blocks the binding of *CD4* (ref. 24). Studies in *LAG3* knockout mice have shown that *LAG3* negatively regulates both the expansion of T-cells and the size of the memory-T-cell pool[25]. During inflammation, both *LAG3* and MHC class II receptors are strongly upregulated, and their interaction may also be involved in the activation of antigen-presenting dendritic cells[26]. In addition, *LAG3* has been shown to confer susceptibility to multiple sclerosis. An analysis of gene expression based on public psoriasis databases[19] indicated that the expression of *CD27* and *LAG3* is upregulated in psoriasis lesions compared with healthy skin samples in both psoriasis cases ($P = 7.47 \times 10^{-15}$ and $P = 6.56 \times 10^{-10}$, respectively) and controls ($P = 4.64 \times 10^{-18}$ and $P = 1.16 \times 10^{-23}$, respectively). A search of the ENCODE database showed that rs2243750 is located within the DHS in medulloblastoma (medullo), osteoblasts (osteobl) and pulmonary artery endothelial cells (HPAECs), suggesting an effect on transcription. Further studies to clarify the identity of the causal gene for psoriasis that is located in the novel susceptibility locus (12p13) identified in this study are therefore warranted.

At 17q12, we identified two psoriasis-associated SNVs, which were in complete LD ($D' = 1$, $r^2 = 1$) (Supplementary Fig. 1c). The 17q12 region has been reported to be associated with several autoimmune diseases, including Crohn's disease, ulcerative colitis, rheumatoid arthritis and systemic lupus erythematosus[27–31]. Within this region, *IKZF3* (*IKAROS* family zinc finger 3) is a plausible candidate gene. *IKZF3* is a member of the IKAROS family of transcription factors involved in preventing the apoptosis of IL2-deprived B-cells[32] and regulating B-cell activation[33]; in addition, it has been implicated in autoimmune disorders, including a lupus-like syndrome that develops in Ikzf3-deficient mice[34]. The expression of *IKZF3* is upregulated in the skin of psoriasis patients, according to public psoriasis databases[19] (PP versus NN: $P = 9.28 \times 10^{-17}$, PP versus PN: $P = 2.25 \times 10^{-16}$), and evidence from the ENCODE database also indicated that rs12936231 is located within the DHS in different cell lines, including mammary epithelial cells (HMECs) and prostate adenocarcinoma cells (LNCaPs), and obtains weak enhancer signal in B-lymphocytes (GM12878).

IFIH1 (Interferon induced with helicase C domain 1) (2q24) encodes a member of the DEAD box proteins family. Members of this family have been implicated in a number of cellular processes. The SNP rs17716942, which is located ~85 kb upstream of *IFIH1*, has been reported to be associated with European population[9]; however it is monomorphic in the Han

Chinese population, which is consistent with data from the 1000 Genomes Project. In this study, we identified a susceptibility intron variant of *IFIH1* (rs13431841), which indicated that *IFIH1* is also a psoriasis susceptibility gene in the Han Chinese population. In addition, *IFIH1* has also been shown to be related to systemic lupus erythematosus[35], type 1 diabetes[36] and Graves' disease[37].

RNF114 (ring finger protein 114) (20q13.13) encodes a 2.4 kb transcript and a 25.7 kDa protein. A previous GWAS identified the SNP rs495337, which is located ~30 kb upstream of *RNF114* and was found to be associated with psoriasis in the European population[11]. This European SNP was observed in our exome sequencing data. In this study, we determined that another SNV (rs4647954) in this LD block is associated with psoriasis in Han Chinese individuals. To evaluate the relationship between rs495337 and rs4647954, we performed LD analysis using exome sequencing data, and we found that rs4647954 identified in our study produced a signal separate from that of the European SNP ($D' = 0.46$, $r^2 = 0.08$). It has been shown that *RNF114* is abundantly expressed in skin, T-lymphocytes and dendritic cells and has a putative role in the regulation of immune responses[11].

For the four SNVs within four known susceptibility loci (*IL12B*, *IFIH1*, *ERAP1* and *RNF114*), the minor alleles identified in the present study were similar between the European and Han Chinese populations, except for rs27043. However, their allele frequencies were highly different, suggesting that these SNVs exhibit allelic heterogeneity within the two ethnic groups (Supplementary Table 3).

In conclusion, we have established three new psoriasis susceptibility loci in a Han Chinese population: *NFKB1*, 12p13.3 and 17q12. Some of these loci have also been implicated in other autoimmune diseases. Furthermore, we confirmed four previously reported psoriasis susceptibility loci (*IL12B*, *IFIH1*, *ERAP1* and *RNF114*) and identified two additional candidate loci (3p21.31 and 17q25). These data, along with data on other reported susceptibility loci, collectively demonstrate the complexity of the heritable contribution to the pathogenesis of psoriasis. Further study will be required if we are to advance our understanding of how the loci identified in this study influence the aetiology of psoriasis.

Methods

Subjects. After the application of quality control procedures to the samples, 15,207 cases and 17,103 healthy controls were enrolled in this study. These cases and controls were Chinese individuals who were enrolled through collaboration among multiple hospitals in China. The samples included in the exome sequencing analysis (the initial stage) were consisted of 781 patients with psoriasis and 676 controls (Table 3), which were mainly selected from previous GWAS samples[12]. A total of 9,946 psoriasis cases and 9,906 controls (Table 3) were recruited for the targeted sequencing analysis (the second stage). The third genotyped cohort consisted of 4,480 psoriasis cases and 6,521 controls (Table 3). The clinical diagnosis of all selected individuals was verified by at least two dermatologists, and clinical information on the patients with psoriasis was collected through a full clinical workup by professional investigators. All the healthy controls were clinically confirmed to be free from psoriasis, autoimmune disorders and systemic disorders and to have a family history of psoriasis and other autoimmune-related disorders (including first-, second- and third-degree relatives). All participants

Table 3 | Summary of the samples analysed in this study.

	Exome sequencing		Targeted sequencing		Genotyping	
	Cases	Controls	Cases	Controls	Cases	Controls
Sample size	781	676	9,946	9,906	4,480	6,521
Mean age (s.d.)	30.92 (12.47)	34.12 (12.92)	35.15 (14.49)	35.81 (16.01)	35.16 (14.80)	29.20 (15.17)
Mean age of onset (s.d.)	22.45 (9.39)		27.02 (13.47)		27.26 (13.44)	
Male/female	458/323	407/269	5,864/4,082	5,329/4,577	2,420/2,060	3,378/3,143

provided written informed consent. The study was approved by the institutional ethical committee of each hospital (The Second Hospital of Anhui Medical University, The First Affiliated Hospital of Anhui Medical University and Huashan Hospital of Fudan University) and was conducted according to the principles of the Declaration of Helsinki.

Exome sequencing. In the first stage, a total of 1,500 exomes were sequenced including 800 cases and 700 controls, at BGI Shenzhen, China. Genomic DNA samples from these individuals were extracted using Flexi Gene DNA kits (Qiagen) and hybridized with a NimbleGen 2.1 M-probe sequence capture array[38] to enrich for exonic DNA in each library. To evaluate exon capture efficiency, the proportions of reads mapping to target regions and to their flanking regions (within 200 bp) were calculated for each individual. Then, we captured and verified the sequencing data for each sample independently using the Illumina Hiseq 2000 platform (San Diego, CA) with an average sequencing depth of $\sim 34 \times$ (ref. 15). Raw image files were processed via the Illumina Pipeline (version 1.3.4) for base calling with the default parameters, and sequences were generated for each individual in the form of 90 bp reads.

Targeted sequencing. In the second stage, we carried out targeted sequencing in a large independent sample of 10,003 cases and 10,002 controls. The target regions consisted of three components: (1) all the exons and exon–intron boundaries for the top 742 genes (based on gene-based association analysis of the exome sequencing data); (2) all the exons and exon–intron boundaries for 622 immune disease-related genes according to the GWAS catalogue (http://www.genome.gov/gwastudies), including 57 candidate genes within 44 susceptibility loci associated with psoriasis; and (3) regions containing the top 133 non-synonymous SNVs obtained through single-variant association analysis of the exome sequencing data (an average of 100 bp per SNV). Furthermore, the proportions of reads that mapped to target regions or their flanking regions (within 200 bp) were calculated for each individual to evaluate the exon capture efficiency. The final target regions, including 1,362 genes, were merged from these three resources (~ 3.2 Mb in total), optimized and used to manufacture a SeqCap EZ Library (Roche NimbleGen, Inc.) to enrich the target regions. Targeted sequencing was finally performed using the Illumina Hiseq 2000 platform with paired-end read lengths of 90 bp, which provided a $64 \times$ coverage on average for these samples[15].

Alignment. We aligned the samples to the NCBI human genome reference assembly (build 36.3) using the Burrows–Wheeler Aligner[39] and processed BAM files to perform re-alignment around known indels using the Genome Analysis ToolKit (GATK v1.6)[40]. All the aligned read data were subjected to Count Covariates (GATK, Count Covariates) on the basis of known SNVs (dbSNPs), and the base quality was subsequently recalibrated (GATK, TableRecalibration).

Sample quality control. The sequencing data for all individuals were evaluated against quality control metrics by verifying (1) an average sequencing depth $\geq 10 \times$ in the first stage (exome sequencing) and $\geq 30 \times$ in the second stage (targeted sequencing); (2) that 90% of the target region was covered $\geq 8 \times$; (3) the GC content; (4) the heterozygosity of SNVs on autosomes; (5) the inbreeding coefficient; (6) population stratification (PCA)[15]; (7) the pairwise IBD; and (8) concordance with the genotyping data. All the QC metrics were reviewed to identify deviations of the data from known or historical norms. Samples that failed QC were not included in further analyses.

A total of 43 individuals from the initial stage and 153 individuals from the second stage failed the sample QC procedure and were removed from further analyses.

Variant calling. We generated genotype information for the target and flanking regions in each individual using SOAPsnp (v1.05)[41], and we further identified and merged candidate SNVs for which at least one individual was called as a high-quality SNV. We then extracted the genotypes of all samples at these positions to build a genotype matrix as an input for the subsequent analysis. A number of filtering criteria were applied to remove false positive calls, and the data quality and error rates were carefully evaluated.

During exome sequencing and targeted sequencing, the detection of variants and genotyping were performed in the target region and the 200 bp flanking region for each individual using SOAPsnp (version 1.05). Then, variation sites such that at least one sample has a confident variation (genotype quality ≥ 20, depth $\geq 8 \times$ and depth of variation allele $\geq 4 \times$) were collected and merged. Finally, the genotype information for all the samples was extracted and combined.

Variant quality control. After the initial SNV calls were generated, we performed further filtering to obtain high-confidence SNVs in the exome sequencing and targeted sequencing steps. For the exome sequencing step, the following criteria were applied: (1) the SNV should be in the target region or the 200 bp flanking region; (2) the SNV should be covered by a total depth of at least 4,500 $\times$ across all individuals (average of 3 $\times$ for each individual); and (3) the SNV should be

observed in at least one individual in such a way that the number of reads containing mutant alleles is >4 in the individual genotype files. For the targeted sequencing data sets, we excluded all SNVs with a call rate below 90% (Q20 and depth $\geq 8 \times$ were considered high quality, others were considered missing). We further excluded SNVs that did not pass the allele balance test (the distributions of the major and minor alleles at the heterozygous positions should be kept in balance, binom test, $P > 1.00 \times 10^{-7}$), the end enrichment test (major and minor alleles should not be enriched at the ends of the reads, Fisher's test, $P > 1.00 \times 10^{-7}$) or the strand bias test (major and minor alleles should not be enriched on either strand, Fisher's test, $P > 1.00 \times 10^{-7}$). Furthermore, SNVs should be located in regions with a homopolymer length ≤ 6 bp, and SNVs located within the MHC and regions of homologous sequence or known indels (indels identified in the 1000 Genomes Project and extending 5 bp) were removed.

We sequenced 24 CHB samples included in the 1000 Genomes Project in this study during the targeted sequencing stage to allow direct comparison of the variants and genotypes called in independent experiments. We relied on the conservative genotype calls provided by the 1000 Genomes Project. There were 35,297 variants showing genotype calls in both studies in the 24 CHB samples. Combining the results for all samples, we estimated an overall concordance rate of 99.59%. In addition, 27 SNVs obtained from the sequencing data were genotyped to determine the fingerprint concordance of the genotype via MALDI-TOF mass spectrometry (Sequenom) in 975 samples, and we observed that 99.49% of the individuals showed a genotype concordance $\geq 99\%$ (ref. 15).

Annotation of SNVs. Using the RefSeq database, we annotated the obtained SNVs into different functional categories with ANNOVAR[42] according to their genetic location and their expected effect on the encoded gene products. In addition, we categorized the SNVs as either known or novel according to their existence in dbSNP (version 135).

Genotyping. In the third stage of the project, we selected the top 22 SNVs with $P < 0.01$ from the targeted sequencing data to be genotyped in an additional 4,480 psoriasis cases and 6,521 controls. Approximately 15 ng of genomic DNA from each sample was used for genotyping. Locus-specific PCR and detection primers were designed using MassARRAY Assay Design 3.0 software (Sequenom) (Supplementary Table 4). After the DNA samples were amplified via multiplex PCR, allele detection was performed through MALDI-TOF mass spectrometry (Sequenom) at the State Key Laboratory of Dermatology, Ministry of Science and Technology, Hefei, Anhui, China. In this case–control cohort, we excluded one SNV showing a call rate of $<90\%$ from further analysis. After quality control, 21 SNVs were analysed in a total of 11,001 individuals at the genotyping replication stage.

Statistical analysis. The single-variant association analysis of SNVs showing an MAF $>0.5\%$ was performed through a standard case/control association analysis (basic χ^2 allelic test) using PLINK 1.07 (ref. 43). We then examined the number of such variants carried by each of the patients and controls, and we compared the number of carriers using Fisher's exact tests or Cochran–Mantel–Haenszel tests[44] across the three collections. Heterogeneity was assessed using the Breslow–Day test. The frequency of alleles in cases and controls, the asymptotic P-values for this test and the odds ratio (OR) were generated in parallel. A Manhattan plot was generated depending on the P-value and location of the SNVs across the genome (Supplementary Fig. 2). Quantile–quantile (QQ) plots were generated using the R package (Supplementary Fig. 3). LD analysis was conducted using Haploview 4.2 (ref. 45).

References

1. Griffiths, C. E. & Barker, J. N. Pathogenesis and clinical features of psoriasis. *Lancet* **370,** 263–271 (2007).
2. Stuart, P. E. *et al.* Genome-wide association analysis identifies three psoriasis susceptibility loci. *Nat. Genet.* **42,** 1000–1004 (2010).
3. Tsoi, L. C. *et al.* Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat. Genet.* **44,** 1341–1348 (2012).
4. Ellinghaus, E. *et al.* Genome-wide meta-analysis of psoriatic arthritis identifies susceptibility locus at REL. *J. Invest. Dermatol.* **132,** 1133–1140 (2012).
5. Liu, Y. *et al.* A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet.* **4,** e1000041 (2008).
6. Ellinghaus, D. *et al.* Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci. *Am. J. Hum. Genet.* **90,** 636–647 (2012).
7. Ellinghaus, E. *et al.* Genome-wide association study identifies a psoriasis susceptibility locus at TRAF3IP2. *Nat. Genet.* **42,** 991–995 (2010).
8. Nair, R. P. *et al.* Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat. Genet.* **41,** 199–204 (2009).
9. Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.* **42,** 985–990 (2010).

10. Sun, L. D. et al. Association analyses identify six new psoriasis susceptibility loci in the Chinese population. Nat. Genet. 42, 1005–1009 (2010).
11. Capon, F. et al. Identification of ZNF313/RNF114 as a novel psoriasis susceptibility gene. Hum. Mol. Genet. 17, 1938–1945 (2008).
12. Zhang, X. J. et al. Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21. Nat. Genet. 41, 205–210 (2009).
13. Huffmeier, U. et al. Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis. Nat. Genet. 42, 996–999 (2010).
14. Zeggini, E. Next-generation association studies for complex traits. Nat. Genet. 43, 287–288 (2011).
15. Tang, H. et al. A large-scale screen for coding variants predisposing to psoriasis. Nat. Genet. 46, 45–50 (2014).
16. Lin, L., DeMartino, G. N. & Greene, W. C. Cotranslational biogenesis of NF-kappaB p50 by the 26S proteasome. Cell 92, 819–828 (1998).
17. Kaufman, C. K. & Fuchs, E. It's got you covered. NF-kappaB in the epidermis. J. Cell Biol. 149, 999–1004 (2000).
18. Kawashima, K. et al. Evaluation of cell death and proliferation in psoriatic epidermis. J. Dermatol. Sci. 35, 207–214 (2004).
19. Gudjonsson, J. E. et al. Assessment of the psoriatic transcriptome in a large sample: additional regulated genes and comparisons with in vitro models. J. Invest. Dermatol. 130, 1829–1840 (2010).
20. Hendriks, J. et al. CD27 is required for generation and long-term maintenance of T cell immunity. Nat. Immunol. 1, 433–440 (2000).
21. Borst, J., Hendriks, J. & Xiao, Y. CD27 and CD70 in T cell and B cell activation. Curr. Opin. Immunol. 17, 275–281 (2005).
22. Agematsu, K. et al. Plasma cell generation from B-lymphocytes via CD27/CD70 interaction. Leuk. Lymphoma 35, 219–225 (1999).
23. O'Shaughnessy, A. & Hendrich, B. CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now. Biochem. Soc. Trans. 41, 777–782 (2013).
24. Bruniquel, D., Borie, N., Hannier, S. & Triebel, F. Regulation of expression of the human lymphocyte activation gene-3 (LAG-3) molecule, a ligand for MHC class II. Immunogenetics 48, 116–124 (1998).
25. Workman, C. J. et al. Lymphocyte activation gene-3 (CD223) regulates the size of the expanding T cell population following antigen activation in vivo. J. Immunol. 172, 5450–5455 (2004).
26. Andreae, S., Piras, F., Burdin, N. & Triebel, F. Maturation and activation of dendritic cells induced by lymphocyte activation gene-3 (CD223). J. Immunol. 168, 3874–3880 (2002).
27. McGovern, D. P. et al. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nat. Genet. 42, 332–337 (2010).
28. Lessard, C. J. et al. Identification of IRF8, TMEM39A, and IKZF3-ZPBP2 as susceptibility loci for systemic lupus erythematosus in a large-scale multiracial replication study. Am. J. Hum. Genet. 90, 648–660 (2012).
29. Franke, A. et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat. Genet. 42, 1118–1125 (2010).
30. Stahl, E. A. et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat. Genet. 42, 508–514 (2010).
31. Anderson, C. A. et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat. Genet. 43, 246–252 (2011).
32. Romero, F., Martinez, A. C., Camonis, J. & Rebollo, A. Aiolos transcription factor controls cell death in T cells by regulating Bcl-2 expression and its cellular localization. EMBO J. 18, 3419–3430 (1999).
33. Wang, J. H. et al. Aiolos regulates B cell activation and maturation to effector state. Immunity 9, 543–553 (1998).
34. Sun, J., Matthias, G., Mihatsch, M. J., Georgopoulos, K. & Matthias, P. Lack of the transcriptional coactivator OBF-1 prevents the development of systemic lupus erythematosus-like phenotypes in Aiolos mutant mice. J. Immunol. 170, 1699–1706 (2003).
35. Cunninghame Graham, D. S. et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. PLoS Genet. 7, e1002341 (2011).
36. Smyth, D. J. et al. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. Nat. Genet. 38, 617–619 (2006).
37. Sutherland, A. et al. Genomic polymorphism at the interferon-induced helicase (IFIH1) locus contributes to Graves' disease susceptibility. J. Clin. Endocrinol. Metab. 92, 3338–3341 (2007).
38. Albert, T. J. et al. Direct selection of human genomic loci by microarray hybridization. Nat. Methods 4, 903–905 (2007).
39. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595 (2010).
40. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).
41. Li, R. et al. SNP detection for massively parallel whole-genome resequencing. Genome Res. 19, 1124–1132 (2009).
42. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010).
43. 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).
44. Wallenstein, S. & Wittes, J. The power of the Mantel-Haenszel test for grouped failure time data. Biometrics 49, 1077–1087 (1993).
45. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265 (2005).

Acknowledgements

Author contributions

X.J.Z conceived this study and obtained financial support. X.J.Z, L.S., L.X., J.W. and Y.L. participated in the study design and were responsible for project management. X.D., D.D., B.L., J.Z., W.Z., H.C., Y.C., F.X., X.F., Z.W., B.L., C.S., H.C., X.G.Z., C.Z., J.X. and W.W. performed sample selection and data management, undertook recruitment, collected phenotype data, undertook related data handling and calculation, managed recruitment and obtained biological samples. Y.S., J.G., X.J., J.X., H.T., Y.L., X.T., X.B.Z. and J.M. designed the bioinformatics and experimental sections, coordinated the data collection, maintained project procedure and performed the data analysis. F.Z., G.C. and X.D.Z. performed the genotyping analysis. H.T., Y.S. and X.B.Z. performed data processing, statistical analysis and bioinformatics investigations. Y.S., J.G., X.J. and J.X. co-wrote the manuscript. All the authors contributed to the final paper, with X.J.Z, L.S., L.X., J.W., S.Y., Y.L., Y.S., J.G., X.J. and J.X. having key roles.

Additional information