ARTICLE

Two new susceptibility loci 1q24.2 and 11p11.2 confer risk to severe acne

Li He[1,*], Wen-Juan Wu[1,*], Jian-Kang Yang[2,3,*], Hui Cheng[4,5], Xian-Bo Zuo[4,5], Wei Lai[6], Tian-Wen Gao[7], Cui-Lin Ma[7], Na Luo[8], Jian-Qing Huang[9], Feng-Yan Lu[10], Ye-Qiang Liu[11], Yi-Jin Huang[12], Qian-Jin Lu[13], Huai-Liang Zhang[14], Lin Wang[15], Wei-Zhen Wang[16], Mei-Mei Wang[17], Sheng-Xiang Xiao[18], Qing Sun[19], Chun-Yang Li[19], Yan-Ping Bai[20], Hui Li[21], Zhan-Chao Zhou[22], Fu-Sheng Zhou[4,5], Gang Chen[4,5], Bo Liang[4,5], Jue Qi[1], Xiao-Yan Yang[1], Ting Yang[1], Xing Zheng[3], Liang-Dan Sun[4,5], Xue-Jun Zhang[4,5] & Ya-Ping Zhang[3]

Severe acne is a chronic inflammatory skin disorder characterized by widespread inflammatory lesions including nodules, cysts and potential scarring. Here we perform the first genome-wide association study of severe acne in a Chinese Han population comprising 1,056 cases and 1,056 controls using the Illumina HumanOmniZhongHua-8 BeadChip. In an independent cohort of 1,860 cases and 3,660 controls of Chinese Han, we replicate 101 SNPs of which 3 showed consistent association. We identify two new susceptibility loci at 11p11.2 (*DDB2*, rs747650, $P_{combined} = 4.41 \times 10^{-9}$ and rs1060573, $P_{combined} = 1.28 \times 10^{-8}$) and 1q24.2 (*SELL*, rs7531806, $P_{combined} = 1.20 \times 10^{-8}$) that are involved in androgen metabolism, inflammation processes and scar formation in severe acne. These results point to new genetic susceptibility factors and suggest several new biological pathways related to severe acne.

[1] Department of Dermatology, First Affiliated Hospital of Kunming Medical University, and Institute of Dermatology & Venereology of Yunnan province, Kunming 650032, China. [2] Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming 650091, China. [3] State Key Laboratory of Genetic Resources and Evolution, and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. [4] Institute of Dermatology and Department of Dermatology at No.1 Hospital, Anhui Medical University, Hefei 230032, China. [5] Key Laboratory of Dermatology (Anhui Medical University), Ministry of Education, Hefei 230032, China. [6] Department of Dermatology, the Third Affiliated Hospital of Sun Yat-sen University, Guangzhou 510000, China. [7] Department of Dermatology, Xijing Hospital, Fourth Military Medical University, Xi'an 710032, China. [8] Department of Dermatology, Southwest Hospital, Third Military Medical University, Chongqing 400038, China. [9] Fuzhou Dermatosis Prevention Hospital, Fuzhou 350025, China. [10] Department of Dermatology, Kunming Medical University Affiliated Hospital Qujing (Qujing First People's Hospital), Qujing 655000, China. [11] Department of Dermatopathology, Shanghai Dermatology Hospital, Shanghai 200443, China. [12] Department of Dermatology, the First Affiliated Hospital of Xiamen University & the First Hospital of Xiamen, Xiamen 361003, China. [13] Department of Dermatology, Second Xiangya Hospital, Central South University, Changsha 410000, China. [14] Institute of Dermatology, Peking Union Medical College and Chinese Academy of Medical Sciences, Nanjing 210000, China. [15] Department of Dermatology, West China Hospital of Sichuan University, Chengdu 610000, China. [16] Department of Dermatology, Wuhan No.1 Hospital, Wuhan 430000, China. [17] Department of Dermatology, First Affiliated Hospital of Baotou Medical College, Baotou 014000, China. [18] Department of Dermatology, Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710000, China. [19] Department of Dermatology, Qilu Hospital of Shandong University, Jinan 250000, China. [20] Department of Dermatology, China-Japan Friendship hospital, Beijing 100000, China. [21] Department of Dermatology, the First Affiliated Hospital of Chongqing Medical University, Chongqing 404100, China. [22] Winona medical cosmetic center, Kunming 650000, China. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to L.H. (email: helikm2662@126.com) or to Y.-P.Z. (email: zhangyp@mail.kiz.ac.cn).

1

Acne vulgaris is a common skin disease characterized by chronic inflammation of the pilosebaceous unit resulting from androgen-induced increased sebum production, altered keratinisation, inflammation and bacterial colonization of hair follicles by *Propionibacterium acnes*[1]. The prevalence of acne varies by age and ethnicity[2–4]. A community-based study performed in China found that acne was present in subjects older than 10 years of age and that the prevalence increased rapidly with age up to 46.8% in the19-year-old group[5]. In subjects with acne, 68.4% had mild (grade I acne of Pillsbury Grade[6]), 26.0% had moderate (grade II and III) and 5.6% had severe acne (grade IV)[5]. Severe acne is characterized by widespread inflammatory lesions, such as nodules, cysts and potential scarring, often creating social handicaps and psychological problems[7].

Increasing evidence from large families and twins implicates genetic factors in the pathogenesis of acne[8,9]. Several candidate genes have thus far been implicated, which include tumour necrosis factor (*TNF*)[10,11], tumour necrosis factor receptor 2 (*TNFR2*)[12], toll-like receptor 2 (*TLR2*)[12], interleukin 1-alpha (*IL-1α*)[13], cytochrome P450 family 1 subfamily A polypeptide 1 (*CYP1A1*)[14], cytochrome P450 family 17 subfamily A polypeptide 1 (*CYP17A1*)[15], cytochrome P450 family 21 subfamily A polypeptide 2 (*CYP21A2*)[16] and androgen receptor (*AR*)[17]. These genes affect two major cellular processes: regulation of steroid hormone metabolism and the innate immune functions of epidermal keratinocytes. However, candidate gene studies were only carried out on small size cohorts, and no previous GWAS of acne has been published. The underlying genetic basis of acne remains poorly understood.

Here we carry out a two-stage GWAS of severe acne involving 2,916 cases and 4,716 controls from a Chinese population to identify susceptibility loci/genes for severe acne. We identify two susceptibility loci at 11p11.2 and 1q24.2 that implicate genes related to androgen metabolism, inflammation processes and scar formation, suggesting their potential involvement in the aetiology of severe acne.

Results

GWAS association results. In the discovery stage, we genotyped 900,015 SNPs using Illumina HumanOmniZhongHua-8 Bead-Chip in 1,056 cases and 1,056 controls of Chinese Han (Table 1). After quality control, 809,305 SNPs (average call rate >98% and MAF >1%) in 1,031 severe acne cases and 1,031 controls of Chinese Han were used in the GWAS discovery analysis.

The scatter plot of *P*-values using logistic regression with adjustment for the gender is shown in Fig. 1. Principal component analysis (PCA) indicated minimal overall inflation of the genome-wide statistical results ($\lambda_{GC} = 1.01$) (Fig. 2, Supplementary Data 1). Moreover, the quantile–quantile plot displayed no global departure from the expected null distribution of *P*-values (Fig. 3). Both of these results indicate negligible

inflation of the genome-wide association signals caused by population stratification, further suggesting that the deviated tail of the *P*-values' distribution reflects some true genetic associations with severe acne. We then carried out logistic regression analysis to assess the genotype–phenotype association.

To perform a fast track replication study, we selected and genotyped 101 SNPs for replication in independent samples of 1,860 cases and 3,660 controls of Chinese Han (Table 1). These SNPs included 86 SNPs with $P < 1 \times 10^{-4}$ in the discovery stage and 15 SNPs with nominal association evidence ($P < 1 \times 10^{-2}$ in the discovery stage) located within or close to nine susceptibility genes that had gene expression profiling evidence for acne or three syndromes with acne symptoms (Apert syndrome, polycystic ovary syndrome and pyogenic arthritis, pyoderma gangrenosum and acne syndrome).

After quality control, 87 SNPs remained for further replication analysis. In the replication stage, three of these SNPs showed consistent association in the independent replication samples and surpassed the threshold for Bonferroni correction in the validation analysis (rs747650, $P = 9.25 \times 10^{-6}$, rs1060573, $P = 2.44 \times 10^{-5}$, rs7531806, $P = 9.22 \times 10^{-7}$). All three SNPs at the two loci showed more significant association in the joint analysis of the combined discovery and replication samples: rs747650 and rs1060573 ($P_{combined} = 4.41 \times 10^{-9}$, OR = 1.24 and $P_{combined} = 1.28 \times 10^{-8}$, OR = 1.23, respectively) at 11p11.2 (*DDB2*) and rs7531806 ($P_{combined} = 1.20 \times 10^{-8}$, OR = 1.22) at 1q24.2 (*SELL*) (Fig. 4, Table 2). The SNPs identified in our GWAS did not situate in previously identified candidate genes, and no previously identified candidate genes were replicated in our study. The association results are shown in Supplementary Data 2.

Analysis of different genetic models. For the three significant SNPs, we used three different genetic models for further analysis: the dominant model, recessive model and additive model. Using a logistic regression analysis, we observed that the association of rs747650 and rs1060573 with severe acne under the additive model ($P = 1.75 \times 10^{-9}$ and $P = 5.66 \times 10^{-9}$, respectively) was more significant than under the dominant ($P = 8.68 \times 10^{-9}$ and $P = 4.55 \times 10^{-8}$, respectively) and recessive models ($P = 2.07 \times 10^{-4}$ and $P = 1.89 \times 10^{-4}$, respectively). Compared with *A/A*, the genotype of homozygote *G/G* (OR$_{Hom}$ = 1.46 and OR$_{Hom}$ = 1.45, respectively) showed larger OR than the heterozygote *G/A* (OR$_{Het}$ = 1.28 and OR$_{Het}$ = 1.25, respectively). The best-fit genetic disease model for rs7531806 was also an additive model ($P = 5.34 \times 10^{-7}$). Compared with *G/G*, the genotype of homozygote *A/A* (OR$_{Hom}$ = 1.40) showed larger OR than the heterozygote *G/A* (OR$_{Het}$ = 1.10) (Table 3).

Gender stratification analysis. These three identified SNPs were selected for stratification analysis by gender to explore the gender-related difference comparing the controls and cases,

Table 1 | Sample characteristics of cases with severe acne and controls.

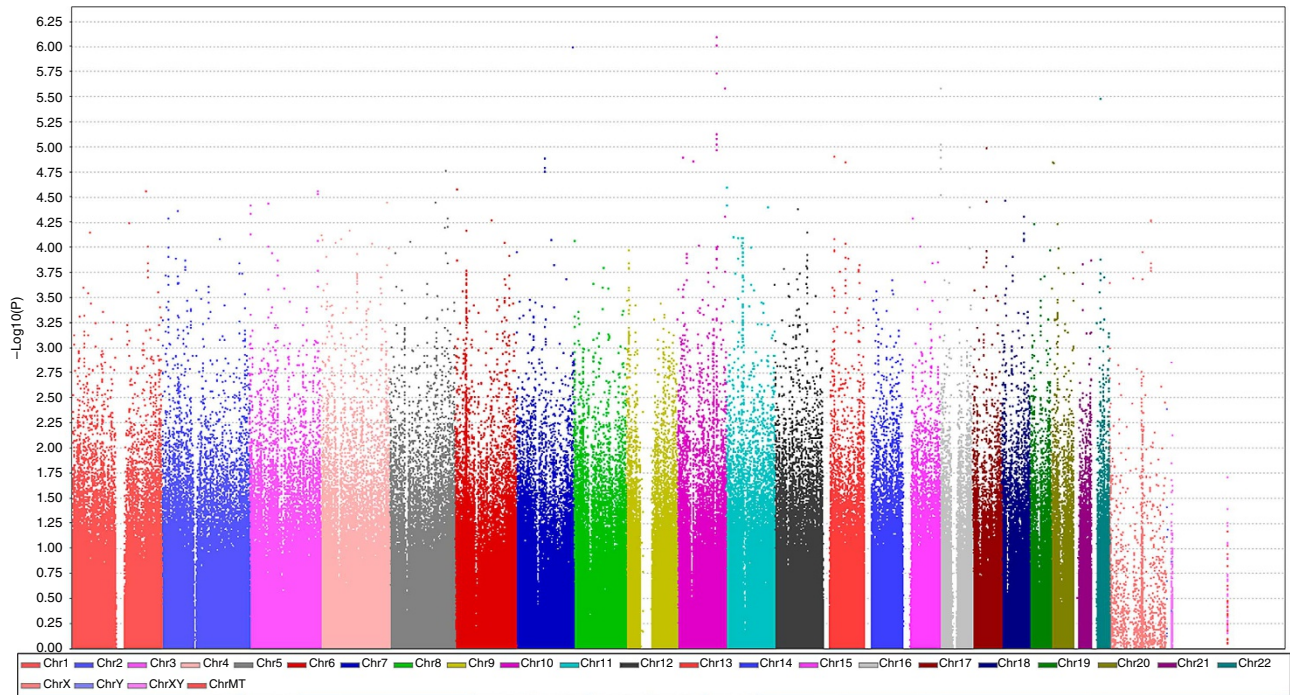|	GWAS		Replication		Combined sample	
	Cases (N = 1,056)	Controls (N = 1,056)	Cases (N = 1,860)	Controls (N = 3,660)	Cases (N = 2,916)	Controls (N = 4,716)
Age*	21.6 ± 5.0	27.7 ± 6.8	21.8 ± 5.1	23.0 ± 7.0	21.7 ± 5.1	24.3 ± 7.3
Age of onset*	16.7 ± 3.4	—	17.0 ± 3.9	—	16.9 ± 3.7	—
Gender						
Male, N (%)	742 (70.3)	661 (62.6)	1,366 (73.4)	1,971 (53.9)	2,108 (72.3)	2,632 (55.8)
Female, N (%)	314 (29.7)	395 (37.4)	494 (26.6)	1,689 (46.1)	808 (27.7)	2,084 (44.2)

*Ages are represented as mean ± s.d.

Figure 1 | Genome-wide association results from the GWAS analysis. The genome-wide *P*-values of the logistic regression test from 809,305 polymorphic SNPs in 1,031 severe acne cases and 1,031 control subjects of Chinese Han descent are presented. The chromosomal distribution of all the *P*-values ($-\log_{10} P$) is shown.
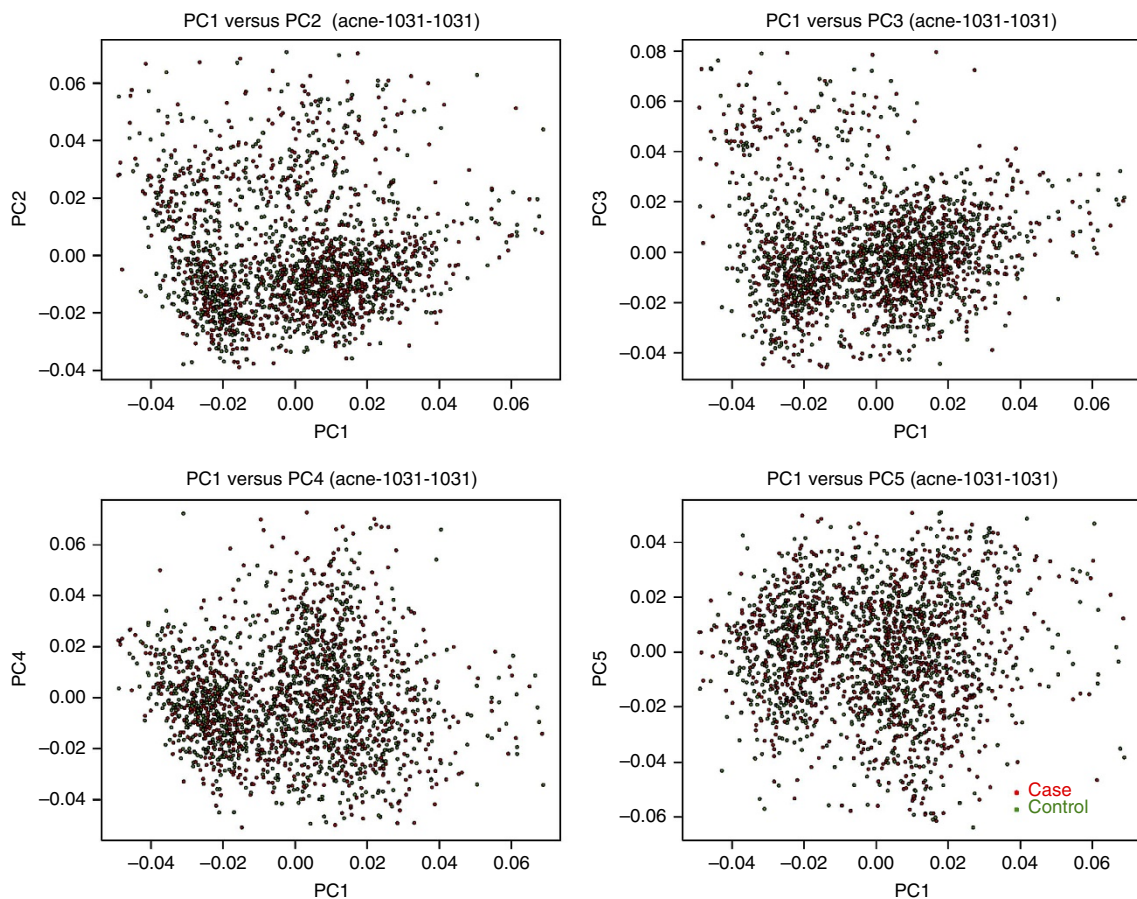

Figure 2 | Principal component analysis of GWAS samples. Principal component analyses (PCA) were performed in our 2,062 GWAS samples (1,031 cases and 1,031 controls). The case–control matching and the low lambda GC values ($\lambda_{GC} = 1.01$) suggested minimal evidence of population stratification.

especially in the cases. However, no nominal heterogeneity was observed among the odds ratios between the male and female after stratification by gender (Table 4).

Discussion

Both SNPs rs747650 and rs1060573 at 11p11.2 highly correlated with one another ($r^2 = 1$) (Fig. 4a). Further analysis using the 1000 genomes data set (http://www.1000genomes.org/) identified another SNP, rs4237547, within the promoter region of *DDB2* (damage-specific DNA binding protein 2) that also highly correlate with rs747650 ($r^2 = 0.97$). The Encyclopedia of DNA Elements (ENCODE, http://genome.ucsc.edu/encode/) data indicated that rs4237547 is located within the DNase I hypersensitivity site and transcription factor binding sites, suggesting a potential effect on gene transcription and regulation. *DDB2* encodes a DNA-binding protein that is the smaller subunit of a
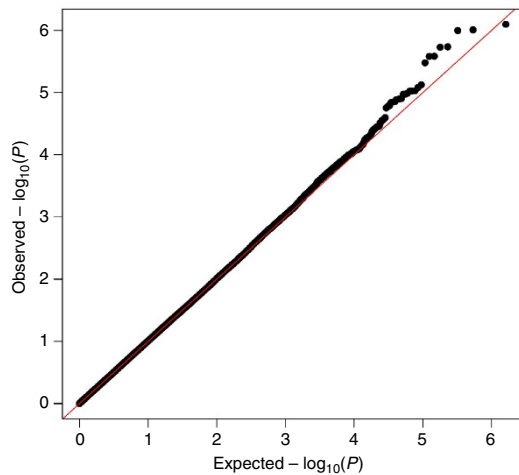
heterodimeric protein complex and participates in nucleotide excision repair. *DDB2* is critical in deciding cell fate (apoptosis or arrest) upon DNA damage[18] and mediates the ubiquitination of the histones H3 and H4 (ref. 19), of which H4 has been shown to be a major component in the antimicrobial action of human sebocytes[20]. *DDB2* has also been identified as a novel androgen receptor-interacting protein, mediating contact with AR and the CUL4A–DDB1 complex for AR ubiquitination/degradation[21]. Consequently, speculating that *DDB2* is a biological candidate gene for severe acne is not unreasonable, particularly given its involvement in androgen metabolism and inflammation processes.

The SNP rs7531806 is located at the loci 1q24.2, covering a gene cluster including *SELL* (selectin L), *SELP* (selectin P) and *SELE* (selectin E) (Fig. 4b). These selectins of adhesion molecules have important roles in regulating homoeostasis and cutaneous inflammation. L-selectin is expressed on the surface of most circulating leukocytes, facilitating leukocyte migration into secondary lymphoid organs and inflammation sites. P- and E-selectin, expressed on inflamed endothelium[22], are likewise responsible for the accumulation of blood leukocytes at sites of inflammation by mediating the adhesion of cells to the vascular lining[23]. Analysis of the gene expression profiles revealed that the mRNA expression of *SELL* was upregulated in acne lesions[24].

L-selectin has been shown to have an important role in mediating cutaneous inflammation by studies using gene-targeted mice[25–27]. Another study also reported a highly significant negative correlation between soluble L-selectin and the inflammatory disorder diffuse systemic sclerosis (dSSc)[28]. Similarly, in mice lacking both L-selectin and intercellular adhesion molecular-1 (ICAM-1) expression, the healing of wounds is delayed, probably due to the decreased leukocyte accumulation into the wound site[29]. Similarly, the absence of both P- and E-selectins markedly reduced recruitment of inflammatory cells and impaired closure of the wounds[30]. Taken together, these findings implicate a potential pathogenic role of selectins in inflammation and the scar-forming processes associated with severe acne.

In summary, we carried out the first GWAS of severe acne among a Chinese population, the findings of which identified new
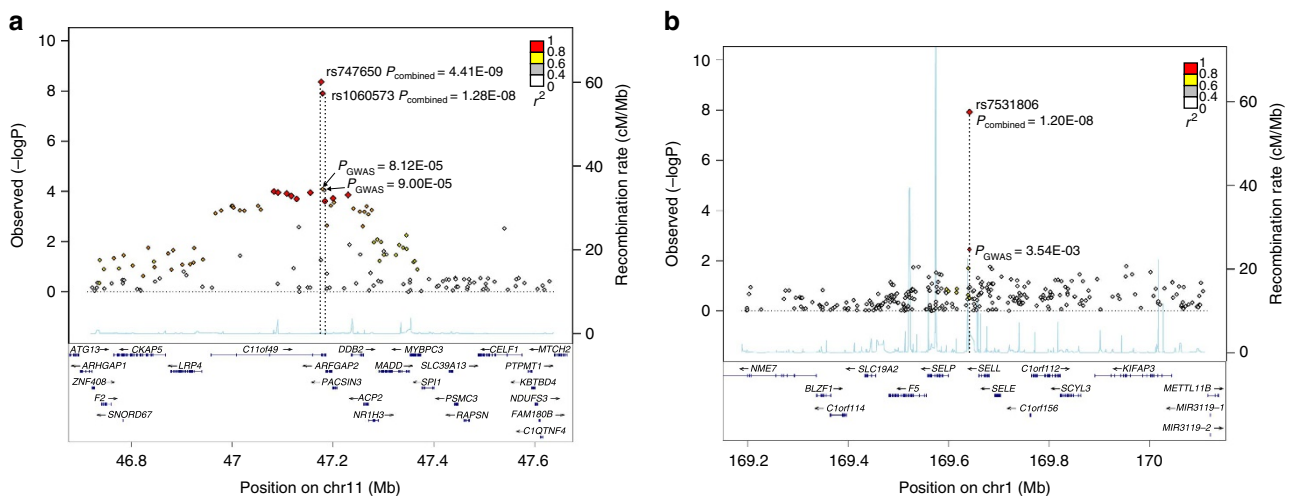

Figure 3 | Quantile–Quantile plots of the observed *P*-values for association in discovery stage. Purple points represent the distribution of logistic regression *P*-values for the association of all the 809,305 SNPs in 1,031 cases and 1,031 control subjects of Chinese Han descent.

Figure 4 | Regional plots of two susceptibility loci for severe acne. (a) 11p11.2 and **(b)** 1q24.2. For each plot, the $-\log_{10}$ *P*-values (left *y*-axis) of SNPs are presented according to their chromosomal positions (*x*-axis). The top genotyped SNP is labeled by rs ID, and the r^2-values of the rest of the SNPs with the top genotyped SNP are indicated by different colours. The genetic recombination rates (estimated using the HapMap Han Chinese in Beijing (CHB) and Japanese in Tokyo (JPT) samples) are represented by light blue lines. Genes within the region are annotated and shown as arrows. *P*-values were generated using logistic regression.

Table 2 | Significant association of three SNPs with severe acne risk.

Gene	SNP	Minor/major allele	Chr.	Position	Stage	MAF Cases	MAF Controls	OR (95% CI)	P-value*	I^{2†}
SELL	rs7531806	A/G	1	169,651,044	GWAS	0.4306	0.3859	1.20 (1.06–1.36)	3.54×10^{-3}	
					Replication	0.4827	0.4314	1.23 (1.13–1.34)	9.22×10^{-7}	
					Combined	0.4627	0.4207	1.22 (1.12–1.28)	1.20×10^{-8}	0
DDB2	rs747650	G/A	11	47,176,005	GWAS	0.3796	0.3210	1.29 (1.14–1.47)	8.12×10^{-5}	
					Replication	0.3696	0.3254	1.22 (1.12–1.33)	9.25×10^{-6}	
					Combined	0.3737	0.3242	1.24 (1.16–1.34)	4.41×10^{-9}	0
DDB2	rs1060573	G/A	11	47,179,829	GWAS	0.3797	0.3215	1.29 (1.14–1.47)	9.00×10^{-5}	
					Replication	0.3692	0.3271	1.20 (1.10–1.31)	2.44×10^{-5}	
					Combined	0.3732	0.3255	1.23 (1.15–1.33)	1.28×10^{-8}	0

Associations are shown for the discovery, replication and combined samples.
*P-values were generated using logistic regression analysis.
$\dagger I^2$ were generated using chi-square (χ^2)-based Cochran's Q statistic

Table 3 | Distribution of genotypes and genetic model analysis for three significant SNPs in combined samples.

	rs747650			rs1060573			rs7531806		
Genotype	AA	GG	GA	AA	GA	GG	GG	AA	GA
Cases	1,050 (39.1%)	362 (13.5%)	1,272 (47.4%)	1,051 (39.1%)	360 (13.4%)	1275 (47.5%)	728 (27.2%)	526 (19.6%)	1,427 (53.2%)
Controls	2,038 (45.7%)	482 (10.8%)	1,937 (43.5%)	2,039 (45.4%)	482 (10.7%)	1973 (43.9%)	1,469 (32.7%)	760 (16.9%)	2,263 (50.4%)
OR (95% CI)	Reference	1.46 (1.25–1.70)	1.28 (1.15–1.41)	Reference	1.45 (1.24–1.69)	1.25 (1.13–1.39)	Reference	1.40 (1.21–1.61)	1.10 (0.96–1.25)
P-value		7.58E−08			2.73E−07			1.89E−06	
Additive		GG/GA/AA			GG/GA/AA			AA/GA/GG		
Cases		362 (13.5%)/ 1,272 (47.4%)/ 1,050 (39.1%)			360 (13.4%)/ 1,275 (47.5%)/ 1,051 (39.1%)			526 (19.6%)/1,427 (53.2%)/728 (27.2%)	
Controls		482 (10.8%)/1,937 (43.5%)/2,038 (45.7%)			482 (10.7%)/1,973 (43.9%)/2,039 (45.4%)			760 (16.9%)/2,263 (50.4%)/1,469 (32.7%)	
OR (95% CI)		1.24 (1.16–1.34)			1.24 (1.15–1.33)			1.20 (1.12–1.28)	
P-value		1.75E−09			5.66E−09			5.34E−07	
Dominant		(GG + GA)/AA			(GG + GA)/AA			(AA + GA)/GG		
Cases		1,634 (60.9%)/ 1,050 (39.1%)			1,635 (60.9%)/ 1,051 (39.1%)			1,953 (72.8%)/ 728 (27.2%)	
Controls		2,419 (54.3%)/ 2,038 (45.7%)			2,455 (54.6%)/ 2,039 (45.4%)			3,023 (67.3%)/ 1,469 (32.7%)	
OR (95% CI)		1.33 (1.21–1.47)			1.31 (1.19–1.45)			1.30 (1.17–1.45)	
P-value		8.68E−09			4.55E−08			8.13E−07	
Recessive		GG/(GA + AA)			GG/(GA + AA)			AA/(GA + GG)		
Cases		362 (13.5%)/ 2,322 (86.5%)			360 (13.4%)/ 2,326 (86.6%)			526 (19.6%)/ 2,155 (80.4%)	
Controls		482 (10.8%)/ 3,975 (89.2%)			482 (10.7%)/ 4,012 (89.3%)			760 (16.9%)/ 3,732 (83.1%)	
OR (95% CI)		1.32 (1.14–1.52)			1.32 (1.14–1.53)			1.21 (1.07–1.37)	
P-value		2.07E-04			1.89E−04			2.45E−03	

P-values were generated using logistic regression analysis.

Table 4 | Association of three SNPs with severe acne analysed by gender stratification in Chinese Han population.

CHR	SNP	Gene	Male versus control P-value†	Male versus control OR (95% CI)	Female versus control P-value†	Female versus control OR (95% CI)	Heterogeneity*	OR for heterogeneity*
1q24.2	rs7531806	SELL	6.57E-06	1.19 (1.10–1.28)	3.91E-03	1.18 (1.05–1.31)	0.843	1.01 (0.90–1.14)
11p11.2	rs747650	DDB2	2.53E-08	1.25 (1.16–1.35)	5.75E-04	1.22 (1.09–1.37)	0.712	1.02 (0.90–1.16)
11p11.2	rs1060573	DDB2	5.22E-08	1.24 (1.15–1.34)	1.18E-03	1.21 (1.08–1.35)	0.646	1.03 (0.91–1.16)

*P-values and odds ratios were calculated using chi-square (χ^2) test from the difference in the odds ratios between males and females.
$\dagger$P-values were combined across discovery and replication stage and generated using logistic regression.

genetic factors that may potentially contribute to severe acne susceptibility. Obviously, further fine-mapping and functional studies are warranted to determine the causal genes within the 1q24.2 and 11p11.2 loci, but the information provided herein should further our basic understanding of the pathogenesis of severe acne, serving as a jumping off point for more sophisticated and focused research efforts.

Methods

Study populations. We performed a two-stage GWAS. The discovery stage included 1,056 severe acne cases and 1,056 controls. The replication stage included 1,860 severe acne cases and 3,660 controls. All samples from both the discovery and replication stages were unrelated individuals of Chinese Han descent, obtained from doctors through collaboration with multiple hospitals within China. To ensure that patient and control inclusion/exclusion were done in a comparable ways across the different centres, we performed three stages of sample gathering, as

follows: in the preparation stage, we created unified epidemiological surveys, which included precise criteria for the definition of a case or a control, and the standard procedures for sample collection, which included information collection, photograph taking and blood collection. Then, in the standard training stage, dermatologists from different centres were required to attend meetings for standard operation training before participant recruitment. Finally, in the sample verification stage, each sample was verified via questionnaires and photographs taken by dermatologist from First Affiliated Hospital of Kunming Medical University before its genomic DNA was extracted. The disease status of each patient and control was also reconfirmed by phone survey.

All the cases of severe acne were diagnosed with grade IV according to the Pillsbury grading system, meaning that patients had many comedones and deep lesions tending to coalesce and canalize, and involving the face and the upper aspects of the trunk[6], without systemic disorders, autoimmune diseases or chronic inflammatory diseases. Clinical information was collected from affected individuals through a full clinical checkup conducted by medical specialists. Additional demographic information was collected from cases and controls through a structured questionnaire. All controls were clinically assessed to be without acne, systemic disorders, autoimmune diseases or family history of severe acne (including first, second and third degree relatives). Given the speculated differences in genetic background between northern and southern Chinese populations[31], the vast majority of the controls and cases for GWAS and replication studies were collected from southern Chinese populations to avoid potential population stratification. Written informed consent was obtained from each subject before sample collections, and this study was approved by the ethical committee of all participating institutions and was done in conformity with the Declaration of Helsinki and subsequent amendments.

Quality control in the discovery stage. The discovery stage was conducted using the Illumina HumanOmniZhongHua-8 BeadChip at the Key Laboratory of Dermatology at Anhui Medical University (Ministry of Education), Hefei, Anhui, China. We performed systematic quality control on the raw genotyping data to filter out both unqualified samples and SNPs. To evaluate the quality of the genotype data for the validation analysis, 100 randomly selected samples from the GWAS stage were re-genotyped using the Sequenom system. The concordance rate between the genotypes from the Illumina HumanOmniZhongHua-8 BeadChip and the Sequenom MassARRAY assay analyses was $>99\%$. Samples with overall call rates of $<98\%$ were excluded from further analysis. Unexpected duplicates or probable relatives were excluded based on pairwise identity by state comparisons using the 'PI_HAT' value in PLINK (all PI_HAT >0.25). After filtering, 1,031 cases and 1,031 controls were retained for analysis.

SNPs were excluded based on three criteria: SNPs with a call rate $<98\%$, with minor allele frequency (MAF) <0.01 in all samples or SNPs with genotype distributions that deviated from those expected by the Hardy–Weinberg equilibrium ($P<1\times10^{-4}$) in the controls. After quality control filtering, 809,305 SNPs remained in the discovery stage.

SNP selection and genotyping in replication stage. SNPs for the replication stage were selected using the following criteria: SNPs with $P<1\times10^{-4}$ for discovery samples or with $P<1\times10^{-4}$ for gender as covariate, which might help reduce the gender effect on the results. A total of 86 SNPs that matched these criteria were accordingly included in the replication stage. We also selected 15 SNPs with nominal association evidence ($P<1\times10^{-2}$ in the discovery stage) that were located within or close to the nine susceptibility genes (*CD14, LIPC, MMP1, SELL, SGK1, TNC, FGFR2, WNT4* and *PSTPIP1*) with gene expression profiling evidence for acne[24] or for three syndromes with acne symptoms (Apert syndrome[32], polycystic ovary syndrome[33] and pyogenic arthritis, pyoderma gangrenosum and acne syndrome[34]). In total, 101 SNPs were selected for the replication. Genotyping analyses of replication were conducted by the Sequenom MassARRAY system at the Key Laboratory of Dermatology at Anhui Medical University (Ministry of Education), Hefei, Anhui, China.

Quality control in replication stage. We excluded SNPs with a call rate $<90\%$ or a deviation from HWE ($P<0.05$) in the controls. After quality control, 87 of the previous 101 SNPs were left for further analysis.

Association analysis in discovery stage. In the discovery stage, we examined potential genetic relatedness based on pairwise identity by state for all of the successfully genotyped samples using PLINK 1.07 software[35]. We used PLINK 1.07 for general statistical analysis. We used the quantile–quantile plots to evaluate the overall significance of the genome-wide association results and the potential effect of population stratification. We further calculated the genomic control inflation factor and found it to be minimal ($\lambda_{GC}=1.01$), suggesting minimal inflation of the genome-wide association results from population stratification. The remaining samples (1,031 cases and 1,031 controls) were subsequently assessed for population outlier and stratification using a principal component analysis (PCA)-based approach. No population outliers were detected. In GWAS stage, single-marker association analyses were performed using logistic regression with gender as a covariate. The Manhattan plot of $-\log_{10}P$ was generated using Haploview v4.2.

Association analysis in replication and combined stage. For the replication studies, 87 SNPs that passed quality control were analysed using logistic regression with gender as a covariate. Joint analysis of all combined samples of Chinese Han was conducted either by random effects model ($I^2>25\%$) or by using fixed-effect model ($I^2<25\%$). In addition to the allelic test of association, the genetic models (dominant, recessive and additive models) were calculated for the associated SNPs. The chi-square (χ^2)-based Cochran's Q statistic was also calculated to test for heterogeneity between groups in stratified analysis. The regional association plot was created using LocusZoom[36]. In stratification analysis by gender, the logistic regression analysis restricted to cases (case-only analyses) was performed with gender as the outcome variable[37].

References

1. Williams, H. C., Dellavalle, R. P. & Garner, S. Acne vulgaris. *Lancet* **379**, 361–372 (2012).
2. Cunliffe, W. J. & Gould, D. J. Prevalence of facial acne vulgaris in late adolescence and in adults. *Br. Med. J.* **1**, 1109–1110 (1979).
3. Freyre, E. A., Rebaza, R. M., Sami, D. A. & Lozada, C. P. The prevalence of facial acne in Peruvian adolescents and its relation to their ethnicity. *J. Adolesc. Health* **22**, 480–484 (1998).
4. Ghodsi, S. Z., Orawa, H. & Zouboulis, C. C. Prevalence, severity, and severity risk factors of acne in high school pupils: a community-based study. *J. Invest. Dermatol.* **129**, 2136–2141 (2009).
5. Shen, Y. *et al.* Prevalence of acne vulgaris in Chinese adolescents and adults: a community-based study of 17,345 subjects in six cities. *Acta Derm. Venereol.* **92**, 40–44 (2012).
6. Witkowski, J. A. & Parish, L. C. The assessment of acne: an evaluation of grading and lesion counting in the measurement of acne. *Clin. Dermatol.* **22**, 394–397 (2004).
7. Goodman, G. J. Acne and acne scarring: why should we treat? *Med. J. Aust.* **171**, 62–63 (1999).
8. Xu, S. X. *et al.* The familial risk of acne vulgaris in Chinese Hans—a case-control study. *J. Eur. Acad. Dermatol. Venereol.* **21**, 602–605 (2007).
9. Herane, M. I. & Ando, I. Acne in infancy and acne genetics. *Dermatology* **206**, 24–28 (2003).
10. Al-Shobaili, H. A., Salem, T. A., Alzolibani, A. A., Robaee, A. A. & Settin, A. A. Tumor necrosis factor-alpha -308 G/A and interleukin 10 -1082 A/G gene polymorphisms in patients with acne vulgaris. *J. Dermatol. Sci.* **68**, 52–55 (2012).
11. Baz, K. *et al.* Association between tumor necrosis factor-alpha gene promoter polymorphism at position -308 and acne in Turkish patients. *Arch. Dermatol. Res.* **300**, 371–376 (2008).
12. Tian, L. M. *et al.* Association study of tumor necrosis factor receptor type 2 M196R and toll-like receptor 2 Arg753Gln polymorphisms with acne vulgaris in a Chinese Han ethnic group. *Dermatology* **221**, 276–284 (2010).
13. Szabo, K. *et al.* Interleukin-1A +4845(G> T) polymorphism is a factor predisposing to acne vulgaris. *Tissue Antigens* **76**, 411–415 (2010).
14. Paraskevaidis, A., Drakoulis, N., Roots, I., Orfanos, C. E. & Zouboulis, C. C. Polymorphisms in the human cytochrome P-450 1A1 gene (CYP1A1) as a factor for developing acne. *Dermatology* **196**, 171–175 (1998).
15. He, L. *et al.* The relationship between CYP17 -34T/C polymorphism and acne in Chinese subjects revealed by sequencing. *Dermatology* **212**, 338–342 (2006).
16. Ostlere, L. S. *et al.* Carrier status for steroid 21-hydroxylase deficiency is only one factor in the variable phenotype of acne. *Clin. Endocrinol.* **48**, 209–215 (1998).
17. Yang, Z. *et al.* Relationship between the CAG repeat polymorphism in the androgen receptor gene and acne in the Han ethnic group. *Dermatology* **218**, 302–306 (2009).
18. Stoyanova, T., Roy, N., Kopanja, D., Bagchi, S. & Raychaudhuri, P. DDB2 decides cell fate following DNA damage. *Proc. Natl Acad. Sci. USA* **106**, 10690–10695 (2009).
19. Wang, H. *et al.* Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol. Cell* **22**, 383–394 (2006).
20. Lee, D. Y. *et al.* Histone H4 is a major component of the antimicrobial action of human sebocytes. *J. Invest. Dermatol.* **129**, 2489–2496 (2009).
21. Chang, S. W. *et al.* DDB2 is a novel AR interacting protein and mediates AR ubiquitination/degradation. *Int. J. Biochem. Cell Biol.* **44**, 1952–1961 (2012).
22. Kansas, G. S. Selectins and their ligands: current concepts and controversies. *Blood* **88**, 3259–3287 (1996).
23. Zarbock, A., Ley, K., McEver, R. P. & Hidalgo, A. Leukocyte ligands for endothelial selectins: specialized glycoconjugates that mediate rolling and signaling under flow. *Blood* **118**, 6743–6751 (2011).
24. Trivedi, N. R., Gilliland, K. L., Zhao, W., Liu, W. & Thiboutot, D. M. Gene array expression profiling in acne lesions reveals marked upregulation of genes involved in inflammation and matrix remodeling. *J. Invest. Dermatol.* **126**, 1071–1079 (2006).

25. Steeber, D. A. *et al.* Leukocyte entry into sites of inflammation requires overlapping interactions between the L-selectin and ICAM-1 pathways. *J. Immunol.* **163,** 2176–2186 (1999).

26. Tedder, T. F., Steeber, D. A. & Pizcueta, P. L-selectin-deficient mice have impaired leukocyte recruitment into inflammatory sites. *J. Exp. Med.* **181,** 2259–2264 (1995).

27. Shimada, Y. *et al.* L-selectin or ICAM-1 deficiency reduces an immediate-type hypersensitivity response by preventing mast cell recruitment in repeated elicitation of contact hypersensitivity. *J. Immunol.* **170,** 4325–4334 (2003).

28. Dunne, J. V., van Eeden, S. F. & Keen, K. J. L-selectin and skin damage in systemic sclerosis. *PLoS One* **7,** e44814 (2012).

29. Nagaoka, T. *et al.* Delayed wound healing in the absence of intercellular adhesion molecule-1 or L-selectin expression. *Am. J. Pathol.* **157,** 237–247 (2000).

30. Subramaniam, M. *et al.* Role of endothelial selectins in wound repair. *Am. J. Pathol.* **150,** 1701–1709 (1997).

31. Xu, S. *et al.* Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am. J. Hum. Genet.* **85,** 762–774 (2009).

32. Melnik, B. C., Vakilzadeh, F., Aslanidis, C. & Schmitz, G. Unilateral segmental acneiform naevus: a model disorder towards understanding fibroblast growth factor receptor 2 function in acne? *Br. J. Dermatol.* **158,** 1397–1399 (2008).

33. Canto, P., Soderlund, D., Carranza-Lira, S. & Mendez, J. P. Mutation analysis of the WNT-4 gene in patients with polycystic ovary syndrome. *Gynecol. Endocrinol.* **22,** 484–487 (2006).

34. Wang, D. *et al.* Inflammation in mice ectopically expressing human Pyogenic Arthritis, Pyoderma Gangrenosum and Acne (PAPA) Syndrome-associated PSTPIP1 A230T mutant proteins. *J. Biol. Chem.* **288,** 4594–4601 (2013).

35. 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).

36. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26,** 2336–2337 (2010).

37. Sun, L. D. *et al.* Association analyses identify six new psoriasis susceptibility loci in the Chinese population. *Nat. Genet.* **42,** 1005–1009 (2010).

Acknowledgements

Author contributions

The study was designed by L.H. and Y.-P.Z. Subject recruitment was undertaken by L.H., W.L, T.-W.G., C.-L.M., N.L., J.-Q.H., F.-Y.L., Y.-Q.L., Y.-J.H., Q.-J.L., H.-L.Z., L.W., W.-Z.W., M.-M.W., S.-X.X., Q.S., C.-Y.L., Y.-P.B., H.L., Z.-C.Z., B.L. and X.-J.Z. Sample selection and preparation was conducted by W.-J.W., J.-K.Y., H.C., J.Q., X.-Y.Y., T.Y. and X.Z.. Sample genotyping was performed by F.-S.Z. and G.C. Statistical analysis was carried out by X.-B.Z., J.-K.Y., W.-J.W., H.C. and L.-D.S. The initial manuscript was drafted by W.-J.W. and J.-K.Y and reviewed by L.H. and Y.-P.Z. The genotyping platform was provided by X.-J.Z.

Additional information