

ARTICLE

Received 22 Dec 2013 | Accepted 1 May 2014 | Published 13 Jun 2014

DOI: 10.1038/ncomms5020

# Genome-wide association study identifies three novel susceptibility loci for severe *Acne vulgaris*

Alexander A. Navarini<sup>1,\*</sup>, Michael A. Simpson<sup>1,\*</sup>, Michael Weale<sup>1,\*</sup>, Jo Knight<sup>2</sup>, Isabelle Carlván<sup>3</sup>, Pascale Reiniche<sup>3</sup>, David A. Burden<sup>4</sup>, Alison Layton<sup>5</sup>, Veronique Bataille<sup>6</sup>, Michael Allen<sup>1</sup>, Robert Pleass<sup>7</sup>, Andrew Pink<sup>1,7</sup>, Daniel Creamer<sup>8</sup>, John English<sup>9</sup>, Stephanie Munn<sup>10</sup>, Shernaz Walton<sup>11</sup>, Carolyn Willis<sup>12</sup>, The Acne Genetic Study Group<sup>†</sup>, Sophie Déret<sup>3</sup>, Johannes J. Voegel<sup>3</sup>, Tim Spector<sup>6</sup>, Catherine H. Smith<sup>7</sup>, Richard C. Trembath<sup>13</sup> & Jonathan N. Barker<sup>1</sup>

*Acne vulgaris* (acne) is a common inflammatory disorder of the cutaneous pilo-sebaceous unit. Here we perform a genome-wide association analysis in the United Kingdom, comparing severe cases of acne ( $n=1,893$ ) with controls ( $n=5,132$ ). In a second stage, we genotype putative-associated loci in a further 2,063 acne cases and 1,970 controls. We identify three genome-wide significant associations: 11q13.1 (rs478304,  $P_{\text{combined}} = 3.23 \times 10^{-11}$ , odds ratio (OR) = 1.20), 5q11.2 (rs38055,  $P_{\text{combined}} = 4.58 \times 10^{-9}$ , OR = 1.17) and 1q41 (rs1159268,  $P_{\text{combined}} = 4.08 \times 10^{-8}$ , OR = 1.17). All three loci contain genes linked to the TGF $\beta$  cell signalling pathway, namely *OVOL1*, *FST* and *TGFB2*. Transcripts of *OVOL1* and *TGFB2* have decreased expression in affected compared with normal skin. Collectively, these data support a key role for dysregulation of TGF $\beta$ -mediated signalling in susceptibility to acne.

<sup>1</sup> Division of Genetics and Molecular Medicine, King's College London, London SE1 9RT, UK. <sup>2</sup> Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON M5T 1R8, Canada. <sup>3</sup> Scientific Division, Galderma R&D, 06410 Sophia-Antipolis, France. <sup>4</sup> Department of Dermatology, Western Infirmary, Glasgow G11 6NT, UK. <sup>5</sup> Department of Dermatology, Harrogate and District Foundation Trust, Harrogate HG2 7SX, UK. <sup>6</sup> Twin Research and Genetic Epidemiology Unit, Division of Genetics and Molecular Medicine, King's College, London SE1 7EH, UK. <sup>7</sup> St John's Institute of Dermatology, Guy's and St. Thomas' NHS Foundation Trust, London SE1 9RT, UK. <sup>8</sup> Department of Dermatology, King's College Hospital NHS Foundation Trust, London SE5 9RS, UK. <sup>9</sup> Department of Dermatology, Queen's Medical Centre, Nottingham University Hospitals NHS Trust, Nottingham NG7 2UJ, UK. <sup>10</sup> Department of Dermatology, Orpington Hospital, King's College Hospital NHS Foundation Trust, Orpington BR6 9JU, UK. <sup>11</sup> Hull and East Yorkshire Hospitals, NHS Trust and Hull York Medical School, Hull HU3 2JZ, UK. <sup>12</sup> Department of Dermatology, Amersham Hospital, Buckinghamshire Healthcare NHS Trust, Amersham HP7 0JD, UK. <sup>13</sup> Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK.

\* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to R.C.T. (email: vp-health@qmul.ac.uk) or to J.N.B. (email: jonathan.barker@kcl.ac.uk) or to C.H.S. (email: catherine.smith@gstt.nhs.uk).

<sup>†</sup>List of members and affiliations appears at the end of the paper.

**A** *cne vulgaris* (acne) is a chronic, inflammatory skin disease characterized by papules, pustules and nodules arising on the face, chest and back. These clinical features reflect underlying sebaceous gland hyperactivity, overgrowth of *Propionibacterium acnes* and inflammation of hair follicles<sup>1</sup>. Typically, the onset of acne occurs during puberty and may persist for many years with irreversible scarring occurring in up to 20% of patients. The consequent social and economic burden is profound with elevated rates of unemployment, depression and low self-esteem<sup>1,2</sup> in those affected.

Family and twin studies indicate a substantial genetic contribution to acne susceptibility with heritability estimates of 78 and 81% (refs 1,3). Genome-wide association studies (GWAS) have made a substantial contribution to the understanding of the genetic basis of several common cutaneous inflammatory disorders<sup>4,5</sup>. However, efforts to characterize the genetic basis of acne have thus far been limited<sup>6,7</sup>. In the Han Chinese population, two genome-wide significant acne susceptibility loci (1q24.2 and 11p11.2) have recently been identified through the study of 1,056 acne cases and 1,056 controls<sup>6</sup>. The modest effect size (odds ratio (OR) of 1.22 and 1.23, respectively) of the reported loci together with the high heritability of acne suggest that additional well-powered GWAS, including studies in different populations, will provide an opportunity to discover further genetic determinants.

Here we capitalized on a national network of dermatologists within the United Kingdom<sup>8</sup> to recruit subjects with severe acne and perform a case-control GWAS leading to the discovery of three new genome-wide significant association signals for acne.

## Results

DNA from 2,001 individuals with severe acne was genotyped on the Illumina HumanOmniExpress-12v1\_H microarray and compared with genotypes from 5,139 UK control subjects previously assayed<sup>9</sup> on the Illumina custom Human1.2M-Duo array. Association analysis was performed with 7.3 million variants that were imputed from the 1000 Genomes Phase I integrated variant set (April 2012)<sup>10</sup> based on genotypes at 386,674 single-nucleotide polymorphisms (SNPs) that had been successfully genotyped in cases and controls. Following quality control and imputation, high-quality data were available for 1,893 cases and 5,132 controls. We performed single SNP analysis using score tests under a logistic regression model including the first four principal components (Supplementary Fig. 1a,b) as covariates. The quantile-quantile plot indicated adequate control for confounders ( $\lambda_{GC} = 1.037$ ; Supplementary Fig. 1c). We found no evidence for association at the acne risk loci reported in the Han Chinese population (1q24.2 (rs7531806,  $P_{\text{discovery}} = 0.597$ ) and 11p11.2 (rs747650,  $P_{\text{discovery}} = 0.337$ ; and rs1060573,  $P_{\text{discovery}} = 0.313$ )). Despite the absence of association at the two reported acne loci, examination of the quantile-quantile plot revealed a modest deviation from the null

distribution, indicating an excess signal of association. We therefore selected 350 SNPs with evidence of association (Methods section; Supplementary Fig. 1d) in the discovery cohort for a second-stage study. We recruited a further 2,207 UK subjects with acne and performed a comparison with 2,087 controls from the Twins UK registry with self-reported absence of a history of acne<sup>3</sup> (Methods section; Supplementary Methods and Supplementary Table 1). Following quality control, high-quality genotypes were available for 286 SNPs across 2,063 cases and 1,970 controls. The second-stage analysis identified 13 SNPs surpassing nominal significance for association ( $P_{\text{second stage}} < 0.05$ ) with the same risk allele as identified in the discovery stage of this study. Of these, five SNPs achieved genome-wide significance in a meta-analysis of the discovery and second-stage cohorts ( $P_{\text{combined}} < 5 \times 10^{-8}$ ; Table 1 and Fig. 1a-c). Three of these SNPs are located within close proximity at 11q13.1 (Fig. 1a). Analysis conditioning on the lead SNP in this region (rs478304,  $P_{\text{combined}} = 3.23 \times 10^{-11}$ , OR = 1.20) reduces the signal at the other two SNPs (rs610037 and rs72922786) to background, implying that a single signal underlies the observed association at these three SNPs. Two further genome-wide significant associations were identified at 5q11.2 (rs38055,  $P_{\text{combined}} = 4.58 \times 10^{-9}$ , OR = 1.17) and 1q41 (rs1159268;  $P_{\text{combined}} = 4.08 \times 10^{-8}$ , OR = 1.17). We found no evidence for interaction between SNPs at these three loci.

We considered genes located within 500 kb of each of the three lead SNPs as potential candidates for a role in acne pathogenesis<sup>5</sup>. In total, these three segments harbour 58 transcripts of which 39 are expressed in skin<sup>11</sup>. *TGFB2* is the closest gene to the association signal at 1q41, located 226 kb from the lead SNP rs1159268. *TGFB2* is a member of the transforming growth factor- $\beta$  (TGF $\beta$ ) family of cytokines, and we noted that genes encoding proteins linked to the TGF $\beta$  cell signalling pathway are located within the regions of association at the other two genome-wide significant loci, namely *OVOL1* (refs 12-14) at 11q13.1 and *FST*<sup>15</sup> at 5q11.2. The observation that a gene related to the TGF $\beta$  pathway is located in each of the three genome-wide significant loci identified in the current study is highly significant ( $P = 6 \times 10^{-5}$ , Methods section). We confirmed the presence of *TGFB2*, *OVOL1* and *FST* transcripts in skin biopsies from individuals with acne and noted a significant reduction in *TGFB2* ( $P_{\text{adjusted}} = 6 \times 10^{-4}$ ) and *OVOL1* ( $P_{\text{adjusted}} = 4 \times 10^{-2}$ ) transcript levels in lesional compared with non-lesional skin (Fig. 2a-c).

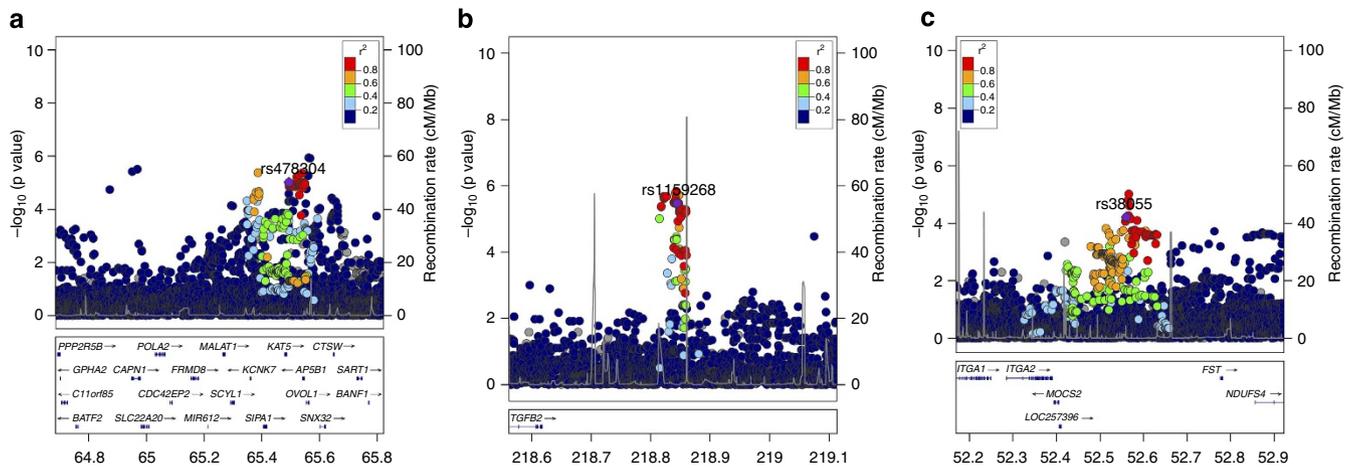
## Discussion

We report three genome-wide significant associated loci with severe acne, each of which contains a gene whose protein product is linked both to the TGF $\beta$  pathway and to skin homeostasis, ranging from keratinocyte proliferation to wound healing<sup>16-18</sup>. The biology of TGF $\beta$  is consistent with a role in acne pathogenesis. Keratinocyte hyperproliferation, which may lead

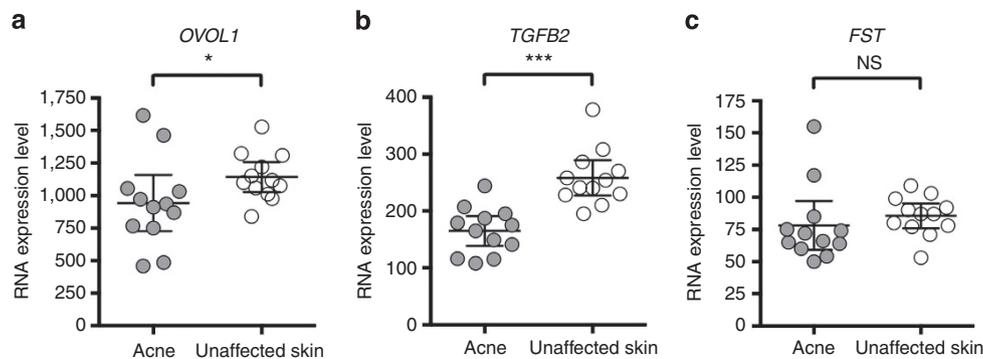
**Table 1 | Loci showing genome-wide significant association with acne.**

Chromosomes	rsID	Position	Gene of Risk interest allele	RAF	Discovery sample		Second-stage sample		Meta-analysis		
					Cases	Controls	$P_{\text{discovery}}$	OR (95% CI)		$P_{\text{second stage}}$	OR (95% CI)
11q13.1	rs478304	65'494'260	<i>OVOL1</i> , others	T	0.60	0.55	$9.58 \times 10^{-6}$	1.20 (1.11-1.29)	$2.65 \times 10^{-7}$	1.26 (1.16-1.38)	$3.23 \times 10^{-11}$
5q11.2	rs38055	52'560'644	<i>FST</i>	A	0.36	0.32	$6.02 \times 10^{-5}$	1.17 (1.08-1.27)	$1.03 \times 10^{-5}$	1.24 (1.13-1.36)	$4.58 \times 10^{-9}$
1q41	rs1159268	218'844'906	<i>TGFB2</i>	A	0.39	0.35	$3.34 \times 10^{-6}$	1.17 (1.08-1.26)	$3.17 \times 10^{-3}$	1.15 (1.05-1.26)	$4.08 \times 10^{-8}$

Abbreviations: CI, confidence interval; RAF, risk allele frequency.



**Figure 1 | Regional association plots at the three genome-wide significant loci.** (a) 11q13.1, (b) 1q41 and (c) 5q11.2. The  $-\log_{10} P_{\text{discovery}}$  values for the SNPs are shown on the left y axis of each plot. SNPs are coloured based on their  $r^2$  with the labelled lead SNP.  $r^2$  is calculated from the 1000 Genomes (March 2012) genotypes. The bottom section of each plot shows the fine-scale recombination rates (right y axis) estimated from individuals in the 1000 Genomes population, and genes are marked by horizontal blue lines. The x axis shows the chromosomal position in Mb.



**Figure 2 | Expression of associated genes in acne.** Normalized expression of associated probe sets for respective genes on Affymetrix RNA expression chips (229396\_at *OVOL1*, 228121\_at *TGFB2* and 207345\_at *FST*) in 12 biopsies of inflammatory acne papules compared with uninvolved skin of the same patients, samples plotted as circles and the overall mean is shown with the 95% confidence interval. Twenty-one of total 39 detected gene transcripts in skin are differentially expressed between lesional and non-lesional skin (Methods section). Stars indicate level of significance of paired *t*-test after Benjamini-Hochberg correction for 38,014 tests, FDR 0.05 (\* $P < 0.05$ , \*\*\* $P < 0.001$ ). NS, not significant.

to follicular obstruction and comedo formation in acne<sup>19,20</sup>, is inhibited by TGFβ<sup>21,22</sup>. Likewise, TGFβ decreases sebaceous gland lipid production<sup>23</sup> which is increased and altered in acne<sup>24,25</sup>. Finally, innate immune responses, elicited during microbial colonization by *P. acnes*<sup>26,27</sup>, may be modulated by TGFβ<sup>28</sup>. Accordingly, transcript levels of *TGFB2* were significantly reduced in fresh inflammatory acne papules in comparison with normal skin (Fig. 2b), as were transcripts of *OVOL1* (Fig. 2a). *OVOL1* is expressed in human<sup>29</sup> and murine<sup>16</sup> skin, in both hair follicles and interfollicular epidermis<sup>16</sup>. It encodes a zinc-finger transcription factor that is a downstream target of TGFβ/BMP7–Smad4 signalling, a pathway known to regulate growth of keratinocytes<sup>12</sup>. TGFβ results in growth arrest of wild-type but not *OVOL1*<sup>-/-</sup> keratinocytes<sup>16,30</sup>, and *OVOL1*<sup>-/-</sup> mice have a thicker epidermis due to hyperproliferation of keratinocytes<sup>16</sup>. A putative role of *OVOL1* in skin pathology is not limited to acne since an association signal at this locus has also been reported in atopic dermatitis<sup>31</sup>. The SNP (rs479844) associated with atopic dermatitis shows no evidence for association with acne in the current study ( $P_{\text{discovery}} = 0.301$ ) nor is it in linkage disequilibrium (LD) with the lead SNP (rs478304) of the acne association signal ( $r^2 = 0.17$ ), suggesting that the association signals are distinct. Our third

genome-wide association signal is at 5q11.2 close to the *FST* gene. Its soluble protein product follistatin binds and neutralizes members of the TGFβ superfamily<sup>15</sup>, including activin, which is involved in scar formation in skin. Accordingly, mice overexpressing follistatin display reduced wound healing<sup>18</sup>, whereas overexpression of activin results in epidermal thickening and dermal fibrosis in normal skin and enhanced granulation tissue formation after wounding<sup>32</sup>.

Our study in European subjects did not detect associations at either of the two previously described acne risk loci<sup>6</sup>. We recognize that our study design, which used an unselected population control cohort in the discovery phase, may impact on the ability to detect true associations. However, the size of the discovery cohort provides sufficient power to detect risk loci with effect sizes of the magnitude reported in the Han Chinese (Supplementary Fig. 2). The disparity may reflect differences in the population-specific genetic determinants<sup>33</sup> and the case and control recruitment criteria of the respective studies. Of interest, the identification of *SELL*, encoding the leukocyte migration protein selectin L<sup>34</sup>, as a candidate gene at one of the loci identified in the Han Chinese population highlights potential inflammatory mechanisms, while the candidate genes reported here relate to tissue remodelling

processes. Both systems are thought to be critical in acne pathophysiology<sup>25,35</sup>.

In summary, we report genome-wide significant associations at three independent genomic loci. This brings the total confirmed regions across the genome known to confer risk for acne to five. Although additional genetic and functional studies are required to clarify the causative alleles at each of the loci, our data support a role of the TGF $\beta$  pathway in predisposition to acne. The current small number of modest effect risk loci taken in the context of the high heritability for acne provides a strong motivation for further genetic analysis, including a search for rare variants, to further our understanding of predisposition to this important disorder.

## Methods

**Clinical samples for genetic study.** The study was designed in accordance with the Declarations of Helsinki, and ethical approval was obtained from the NRES Committee London-Westminster (reference CLRN 05/Q0702/114). 17 UK clinical centres recruited and obtained informed consent from individuals with acne for the discovery stage study, and 39 UK centres for the second-stage study. Clinical assessment was undertaken by trained dermatologists using one or more of the following criteria for a diagnosis of severe acne: (a) nodulocystic disease; (b)  $\geq 5$  points in any body region assessed by the validated Leeds clinical acne score that uses a colour photographic acne grading scheme to evaluate the severity of involvement of body regions (face 0–12, chest 0–8 and back 0–8)<sup>36</sup>; (c) requiring treatment with isotretinoin; and (d) presence of rare and severe forms of acne such as, for example, *acne fulminans*. Inclusion and exclusion criteria are detailed in the Supplementary Methods, and were the same for the discovery and second-stage cohorts that comprised 2,001 and 2,207 individuals with severe acne, respectively. DNA for the case cohorts was extracted from saliva samples collected with Oragene kits ( $n = 3,541$ ) or from whole blood ( $n = 667$ ). The controls for the discovery-stage cohort comprised 2,478 healthy blood donors from the United Kingdom Blood Service (UKBS) collection and 2,661 individuals from the 1958 Birth Cohort (58C) data set<sup>9</sup> (<http://www.wtccc.org.uk/ccc2/>). In the second-stage study, 2,087 controls were selected from the Twins UK registry and were enriched (87%) for individuals with no history of acne based on self-completed questionnaire.

**Clinical samples for expression study.** Paired skin biopsies of lesional and non-lesional skin were obtained from the upper back of acne patients of European origin (11 males and 1 female). The mean age of patients was 23.5 years ( $\pm 3.6$  years s.d.). All were untreated and had inflammatory acne with evidence of scarring. Acute inflammatory papules, defined as having developed within 48 h, were biopsied.

**Genotyping and quality control in discovery stage.** Genome-wide genotyping for the discovery case cohort was conducted by Illumina Corp. (CA) on Illumina HumanOmniExpress-12v1\_H microarrays. Bead intensity data were processed and normalized for each sample in BeadStudio, data for successfully genotyped samples were extracted and genotypes were called using Illuminus (Illumina Corp.). Controls for the discovery stage had been previously genotyped on an Illumina 1.2M platform at the Sanger Institute, Cambridge<sup>9</sup>. Non-overlapping SNPs (255,193) that were only assessed in either cases or controls were excluded during merging of case and control data sets. Using the Plink software package, we followed accepted quality control procedures<sup>37</sup> and excluded 108 cases and 7 controls based on the following criteria:  $> 2\%$  missing genotypes, discrepancy of inferred gender with sample information, cryptic relatedness (PI\_HAT  $> 0.2$ , one of each pair of related individuals were excluded) and excessive heterozygosity. Population stratification was assessed using Principal Component Analysis (PCA) implemented in EIGENSOFT4.2 (refs 38,39) using an LD pruned ( $r^2 < 0.2$ ) set of 93,906 autosomal SNPs that are not located within regions of long-range-LD<sup>40</sup>. Plots of PCA axis 1 and 2 identified 83 outlier samples (63 cases and 20 controls) who were excluded from further analysis and PCA was repeated. SNPs with a call rate  $< 99\%$ , Hardy–Weinberg equilibrium test  $P < 10^{-6}$ , MAF  $< 0.5\%$  or significantly different call rates between cases and controls ( $P < 0.05$ ), and sex-chromosomal SNPs were removed from further analysis. After quality control, genotypes for 386,674 SNPs in 1,893 cases and 5,132 controls remained.

**Imputation and association testing in discovery stage.** Imputation was performed in a two-stage approach. Initially, the data were phased with ShapIT<sup>41</sup> (<http://www.shapeit.fr>) and subsequently imputed against the 1000 Genomes reference panel<sup>10</sup> with IMPUTE2 (ref. 42; [http://mathgen.stats.ox.ac.uk/impute/impute\\_v2.html](http://mathgen.stats.ox.ac.uk/impute/impute_v2.html)). SNPs imputed with relatively high confidence (estimated  $r^2$  between imputed and true genotypes  $> 0.5$ ) were included in the analysis. SNPTEST v2.4.0 (ref. 43); ([https://mathgen.stats.ox.ac.uk/genetics\\_software/snpctest/snpctest.html](https://mathgen.stats.ox.ac.uk/genetics_software/snpctest/snpctest.html)) was used for association testing using the first four covariates (Supplementary Fig. 1a,b) to correct for population stratification and other potential confounders. The over dispersion factor of association test statistics  $\lambda$  was

1.037, suggesting that the impact of systematic differences between cases and controls were within acceptable limits (Supplementary Fig. 1c,d).

**SNP selection for second-stage study.** In total, 350 SNPs were selected for the second-stage study (Supplementary Data 1), 108 lead SNPs with a  $P_{\text{discovery}} < 1 \times 10^{-5}$  that represent the strongest evidence of statistical association at each respective locus. A further 242 SNPs with a  $P_{\text{discovery}} < 1.3 \times 10^{-3}$  were selected as either proxies for the lead SNPs or located in close proximity to genes highly expressed in skin (*SDC1* and *TYRP1*), published candidate genes (*CYP21A2* (ref. 44) or other putative roles in tissue remodelling (*FGF2* and *FGF22*) or the immune response (*IL1R1/2*, *IL1RL1/2*, *IL18R1*, *ICAM4* and *CFB*). We preferentially selected genotyped over imputed SNPs and SNPs with a minor allele frequency of  $> 0.02$ , including manual inspection of cluster plots of genotyped SNPs.

**Genotyping, quality control and association testing in second stage.** Genotyping was undertaken at the KCL/GSTT BRC Genomics Core Facility using the Illumina GoldenGate assay (GT-221-1109-GGGT Universal-32). The generated data were subject to sample and SNP quality control procedures as described above, including visual inspection of cluster plots. We fitted a logistic regression model for case or control status with no covariates at each SNP and a single parameter for the genetic effect (a multiplicative effect on the risk scale and an additive effect on the log-odds scale). The inflation was  $\lambda = 1.047$ . Of the 350 SNPs selected for the second stage, 286 passed quality control and genotypes were available for 2,063 cases and 1,970 controls (Supplementary Data 1). Of these, 18 achieved nominal significance ( $P_{\text{second stage}} < 0.05$ ) of which 13 (4.5%) were directionally consistent with the same risk allele as was observed in the discovery study. Seven of the 108 second-stage SNPs passing quality control and whose selection was based solely on statistical criteria ( $P_{\text{discovery}} < 10^{-5}$ ) achieved nominal significance (six directionally consistent with the discovery study). The controls of the second-stage study were predominantly female (Supplementary Table 1). As our analyses were limited to autosomal regions of the genome, Mendelian segregation ensures that the analysed genotypes and gender are uncorrelated. We tested to establish whether the genetic effects were modulated by gender (modelling SNP allele dosage + gender + gender  $\times$  dosage + the first four principal components as covariates), which failed to demonstrate an interaction for any of the genome-wide significant SNPs ( $P$ -values for interaction term: rs38055,  $P = 0.57$ ; rs478304,  $P = 0.22$ ; and rs1159268  $P = 0.99$ ). Researchers can request access to the data via email to the corresponding authors or via the relevant data access committee (<http://www.ebi.ac.uk/ega>; <http://www.twinsuk.ac.uk/data-access/>).

**Regional plots and identification of genes within associated regions.** Associated regions were plotted using the stand-alone version of LocusZoom<sup>45</sup> using the  $P_{\text{discovery}}$  values from the discovery study and LD information from the 1000 Genomes Project. Regions of association were defined by a 500-kb interval flanking each association signal<sup>5</sup>.

**Evaluation of the observation of TGF $\beta$ -associated genes within the genome-wide association signals.** Overlapping 500 kb regions flanking the SNPs passing quality control in the second stage were merged, resulting in 180 distinct loci of 1.06 Mb on average (s.d. 0.3 Mb). The genes in the 177 loci interrogated in the second stage in which we did not observe genome-wide signals were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) TGF $\beta$  pathway, identifying five genes (*MYC*, *SMAD1*, *TNF*, *MAPK3* and *GDF7*). A hypergeometric exact test<sup>46</sup> was used to evaluate the probability of the observation of three genes linked to the TGF $\beta$  pathway located in three genome-wide significant loci when a total of eight such genes are present across all 180 loci.

**Conditional analysis.** We observed three SNPs (rs478304, rs610037 and rs72922786) with genome-wide significance at the 11q13.1 locus. We performed conditional analysis on the second-stage data set, conditioning the analysis of rs610037 and rs72922786 with the observed genotypes of the lead SNP rs478304.

**Skin expression studies.** Punch biopsies were immediately transferred to RNA-later Tissue Protect Tubes (Qiagen). Total RNA was extracted using miRNeasy extraction kits (Qiagen) including DNase I treatment (27U, 15 min). All RNA samples were evaluated by NanoDrop 8800 and displayed a good profile with a RNA Integrity Number  $> 7$  and a concentration  $> 60 \text{ ng } \mu\text{l}^{-1}$ . cRNA labelling and hybridization were performed according to Affymetrix protocols. Biotinylated RNA and fragmented RNA were assessed for relative length on an Agilent 2100 bioanalyzer. Fragmented RNA was hybridized on Affymetrix U133 Plus 2.0 chips. Affymetrix RMA (Robust Multichip Average) software was used to perform the quality control of the study before analysing the data, namely normalized using the RMA method. RMA consists of three steps: background adjustment, quantile normalization and finally summarization. Genes within a 500-kb interval (see above) of the association signals were assessed for transcript abundance and differential expression between lesional and non-lesional skin. Statistical analysis was performed with paired two-tailed  $t$ -test adjusted for 38,014 independent tests.

## References

- Williams, H. C., Dellavalle, R. P. & Garner, S. Acne vulgaris. *Lancet* **379**, 361–372 (2012).
- Cunliffe, W. J. Acne and unemployment. *Br. J. Dermatol.* **115**, 386 (1986).
- Bataille, V., Snieder, H., MacGregor, A. J., Sasieni, P. & Spector, T. D. The influence of genetics and environmental factors in the pathogenesis of acne: a twin study of acne in women. *J. Invest. Dermatol.* **119**, 1317–1322 (2002).
- 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).
- Tsoi, L. C. *et al.* Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat. Genet.* **44**, 1341–1348 (2012).
- He, L. *et al.* Two new susceptibility loci 1q24.2 and 11p11.2 confer risk to severe acne. *Nat. Commun.* **5**, 2870 (2014).
- Zhang, M., Qureshi, A. A., Hunter, D. J. & Han, J. A genome-wide association study of severe teenage acne in European Americans. *Hum. Genet.* **133**, 259–264 (2013).
- NIHR. Clinical Research Network (2014) [http://www.nihr.ac.uk/infrastructure/Pages/infrastructure\\_clinical\\_research\\_networks.aspx](http://www.nihr.ac.uk/infrastructure/Pages/infrastructure_clinical_research_networks.aspx).
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
- Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- Petryszak, R. *et al.* Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res.* **42**, D926–D932 (2014).
- Kowanetz, M., Valcourt, U., Bergstrom, R., Heldin, C. H. & Moustakas, A. Id2 and Id3 define the potency of cell proliferation and differentiation responses to transforming growth factor beta and bone morphogenetic protein. *Mol. Cell. Biol.* **24**, 4241–4254 (2004).
- Descargues, P., Sil, A. K. & Karin, M. IKKalpha, a critical regulator of epidermal differentiation and a suppressor of skin cancer. *EMBO J.* **27**, 2639–2647 (2008).
- Gomis, R. R. *et al.* A FoxO-Smad synexpression group in human keratinocytes. *Proc. Natl Acad. Sci. USA.* **103**, 12747–12752 (2006).
- Massague, J. & Chen, Y. G. Controlling TGF-beta signaling. *Genes Dev.* **14**, 627–644 (2000).
- Nair, M. *et al.* Ovol1 regulates the growth arrest of embryonic epidermal progenitor cells and represses c-myc transcription. *J. Cell Biol.* **173**, 253–264 (2006).
- Choi, Y. & Fuchs, E. TGF-beta and retinoic acid: regulators of growth and modifiers of differentiation in human epidermal cells. *Cell. Regul.* **1**, 791–809 (1990).
- Wankell, M. *et al.* Impaired wound healing in transgenic mice overexpressing the activin antagonist follistatin in the epidermis. *EMBO J.* **20**, 5361–5372 (2001).
- Knaggs, H. E., Holland, D. B., Morris, C., Wood, E. J. & Cunliffe, W. J. Quantification of cellular proliferation in acne using the monoclonal antibody Ki-67. *J. Invest. Dermatol.* **102**, 89–92 (1994).
- Plewig, G., Fulton, J. E. & Kligman, A. M. Cellular dynamics of comedo formation in acne vulgaris. *Arch. Dermatol. Forsch.* **242**, 12–29 (1971).
- Pietropol, J. A., Holt, J. T., Stein, R. W. & Moses, H. L. Transforming growth factor beta 1 suppression of c-myc gene transcription: role in inhibition of keratinocyte proliferation. *Proc. Natl Acad. Sci. USA.* **87**, 3758–3762 (1990).
- Buschke, S. *et al.* A decisive function of transforming growth factor-beta/Smad signaling in tissue morphogenesis and differentiation of human HaCaT keratinocytes. *Mol. Biol. Cell.* **22**, 782–794 (2011).
- McNairn, A. J. *et al.* TGFbeta signaling regulates lipogenesis in human sebaceous glands cells. *BMC. Dermatol.* **13**, 2 (2013).
- Thiboutot, D. *et al.* Activity of the type 1 5 alpha-reductase exhibits regional differences in isolated sebaceous glands and whole skin. *J. Invest. Dermatol.* **105**, 209–214 (1995).
- Zouboulis, C. C. *et al.* What is the pathogenesis of acne? *Exp. Dermatol.* **14**, 143–152 (2005).
- Kim, J. *et al.* Activation of toll-like receptor 2 in acne triggers inflammatory cytokine responses. *J. Immunol.* **169**, 1535–1541 (2002).
- Kistowska, M. *et al.* IL-1beta Drives Inflammatory Responses to Propionibacterium acnes In Vitro and In Vivo. *J. Invest. Dermatol.* **134**, 677–685 (2013).
- Sanjabi, S., Zenewicz, L. A., Kamanaka, M. & Flavell, R. A. Anti-inflammatory and pro-inflammatory roles of TGF-beta, IL-10, and IL-22 in immunity and autoimmunity. *Curr. Opin. Pharmacol.* **9**, 447–453 (2009).
- Rustici, G. *et al.* ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res.* **41**, D987–D990 (2013).
- Koster, M. I. & Roop, D. R. Mechanisms regulating epithelial stratification. *Annu. Rev. Cell. Dev. Biol.* **23**, 93–113 (2007).
- Paternoster, L. *et al.* Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat. Genet.* **44**, 187–192 (2012).
- Bamberger, C. *et al.* Activin controls skin morphogenesis and wound repair predominantly via stromal cells and in a concentration-dependent manner via keratinocytes. *Am. J. Pathol.* **167**, 733–747 (2005).
- Williams, A. L. *et al.* Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* **506**, 97–101 (2014).
- Wedepohl, S. *et al.* L-selectin—a dynamic regulator of leukocyte migration. *Eur. J. Cell. Biol.* **91**, 257–264 (2012).
- 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).
- O'Brien, S., Lewis, J. & Cunliffe, W. The Leeds revised acne grading system. *J. Dermatol. Treat.* **9**, 215–220 (1998).
- Weale, M. E. Quality control for genome-wide association studies. *Methods Mol. Biol.* **628**, 341–372 (2010).
- Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
- Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
- Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135 (author reply 135–139) (2008).
- Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
- Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
- Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
- Caputo, V., Fiorella, S., Curiale, S., Caputo, A. & Niceta, M. Refractory acne and 21-hydroxylase deficiency in a selected group of female patients. *Dermatology* **220**, 121–127 (2010).
- Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
- Rice, J. A. in *Mathematical Statistics and Data Analysis* 3rd edn (Duxbury Press, 2007).

## Acknowledgements

We thank Saxena Alka (London, UK); Ulrike Blume-Peytavi (Berlin, Germany); Leaca Crawford<sup>34</sup>; Jana Estafan<sup>24</sup>; Darren Geoghegan<sup>1</sup>; Dan Glass<sup>7</sup>; Alison Gosh<sup>35</sup>; Naomi Hare<sup>1</sup>; Helen Holmes<sup>1</sup>; Eleanor Mallon<sup>35</sup>; Karen Markwell (West Hertfordshire, UK); Philippe Martel<sup>4</sup>; Carine Marty<sup>4</sup>; Corinne Ménigot<sup>4</sup>; Efterpi Papouli (London, UK); Anne Thompson<sup>51</sup>; Kate Thornberry<sup>1</sup>; Bianca Tobin<sup>46</sup>; and the participating patients and supporting staff in all the study centres. This study was supported in part by a research grant from Galderma SA to J.N.B., R.C.T., C.H.S., M.S., M.W., J.K., T.S.; by a British Skin Foundation Grant to C.H.S.; by Stiftung Dr Max Cloëtta; UNISCIENTIA/SNF PASM3\_140074/W.G. Siegenthaler grant to A.A.N.; and by an Medical Research Council/British Association of Dermatologists/British Skin Foundation Clinical Research Training Fellowship to A.P. The research was funded/supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the National Health Service (NHS), the National Institute for Health Research (NIHR) or the Department of Health. We acknowledge the support of the NIHR, through the Dermatology Clinical Research Network, with case ascertainment.

## Author contributions

A.A.N., M.A.S., M.W., J.K., I.C., P.R., S.D., J.J.V. performed experiments and analysed data; A.A.N., M.A.S., M.W., R.C.T., C.H.S. and J.N.B. wrote the paper; C.H.S., R.C.T. and J.J.B. supervised the project; D.A.B., A.L., V.B., D.C., J.E., S.M., S.W., C.W., T.S. and the Acne Genetic Study Group contributed clinical information and samples; M.A., R.P., A.P. gave technical support.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://www.nature.com/reprints/index.html>.

**How to cite this article:** Navarini, A.A. *et al.* Genome-wide association study identifies three novel susceptibility loci for severe *Acne vulgaris*. *Nat. Commun.* **5**:4020 doi: 10.1038/ncomms5020 (2014).

Victoria Akhras<sup>16</sup>, Anton Alexandroff<sup>17</sup>, Alex Anstey<sup>18</sup>, Jaskiran Azad<sup>19</sup>, Omar Aziz<sup>20</sup>, David Baudry<sup>1</sup>, Nigel Burrows<sup>21</sup>, Aamir Butt<sup>22</sup>, Peter Cartwright<sup>23</sup>, Anna Chapman<sup>24</sup>, Tim Clayton<sup>25</sup>, Sandeep Cliff<sup>26</sup>, William Cunliffe<sup>27</sup>, Tim Cutler<sup>20</sup>, Brigid Daly<sup>28</sup>, Eleanor Damian<sup>1</sup>, Amrit Darvay<sup>29</sup>, Claudia DeGiovanni<sup>30</sup>, Anthony Downs<sup>31</sup>, Michael Duckworth<sup>1</sup>, Colm Dwyer<sup>32</sup>, Adam Ferguson<sup>33</sup>, Colin Fleming<sup>34</sup>, Zoe Fox<sup>1</sup>, Elizabeth Frazer-Andrews<sup>35</sup>, Justo Gonzalez<sup>1</sup>, Mark Goodfield<sup>36</sup>, Robert Graham<sup>37</sup>, Clive Grattan<sup>38</sup>, Neil Green<sup>1</sup>, Simon Hazelwood-Smith<sup>1</sup>, Harmut Hempel<sup>39</sup>, Sue Hood<sup>20</sup>, Bronwyn Hughes<sup>40</sup>, Effie Ladoyanni<sup>41</sup>, Calum Lyon<sup>42, 43</sup>, Ali Mahmud<sup>44</sup>, Mohsin Malik<sup>45</sup>, Simon Meggitt<sup>46</sup>, Andrew Messenger<sup>47</sup>, Yaaseen Moosa<sup>16</sup>, Anthony Ormerod<sup>48</sup>, Varsha Patel<sup>1</sup>, Deepak Rallan<sup>20</sup>, Sue Robertson<sup>35</sup>, Janet Ross<sup>49</sup>, Ingrid Salvary<sup>37</sup>, Rachel Wachsmuth<sup>50</sup>, Shyamal Wahie<sup>51</sup> & Diane Williamson<sup>52</sup>.

<sup>16</sup>St George's Healthcare NHS Trust, St George's, UK. <sup>17</sup>Leicester Royal Infirmary, University Hospitals of Leicester NHS Trust, Leicester, UK. <sup>18</sup>St Woolos Hospital, Aneurin Bevan Health Board, Gwent, Gwent, UK. <sup>19</sup>James Cook University Hospital, South Tees Hospitals NHS Foundation, South Tees, UK. <sup>21</sup>Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. <sup>22</sup>Scunthorpe General Hospital, Northern Lincolnshire and Goole Hospitals NHS Foundation Trust, Northern Lincolnshire, UK. <sup>23</sup>Queen's Hospital, Burton Hospitals NHS Foundation Trust, Burton, UK. <sup>24</sup>Queen Elizabeth Hospital Woolwich, Lewisham and Greenwich NHS Trust, Woolwich, UK. <sup>25</sup>Salford Royal NHS Foundation Trust, Salford, Salford, UK. <sup>26</sup>East Surrey Hospital, Surrey and Sussex Healthcare NHS Trust, Redhill, UK. <sup>27</sup>Leeds Teaching Hospitals NHS Trust, Leeds, UK. <sup>28</sup>Royal Blackburn Hospital, East Lancashire NHS Trust, Blackburn, UK. <sup>29</sup>Southmead Hospital, North Bristol NHS Trust, Bristol, UK. <sup>30</sup>Brighton General Hospital, Brighton and Sussex University Hospitals NHS Trust, Brighton, UK. <sup>31</sup>Royal Devon & Exeter Hospital, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK. <sup>32</sup>Crosshouse Hospital, NHS Ayrshire and Arran, Ayrshire & Arran, UK. <sup>33</sup>Royal Derby Hospitals, Derby Hospitals NHS Foundation Trust, Derby, UK. <sup>34</sup>Ninewells Hospital and Medical School, NHS Tayside, Dundee, Dundee, UK. <sup>36</sup>Chapel Allerton Hospital, The Leeds Teaching Hospitals NHS Trust, Leeds, UK. <sup>37</sup>James Paget University Hospital NHS Foundation Trust, Great Yarmouth, Gt.Yarmouth, UK. <sup>38</sup>Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK. <sup>39</sup>Great Western Hospitals NHS Foundation Trust, Swindon, UK. <sup>20</sup>Ipswich Hospital NHS Trust, Ipswich, UK. <sup>40</sup>St Mary's Hospital, Portsmouth Hospitals NHS Trust, Portsmouth, UK. <sup>41</sup>Corbett Hospital, The Dudley Group of Hospitals NHS Foundation Trust, Dudley, UK. <sup>42</sup>Scarborough Hospital, Hull and East Yorkshire Hospitals NHS Trust, Scarborough, UK. <sup>43</sup>York Hospitals NHS Foundation Trust, York, UK. <sup>44</sup>Eastbourne District General Hospital, East Sussex Healthcare NHS Trust, Eastbourne, UK. <sup>45</sup>Broomfield Hospital, Mid Essex Hospital Services NHS Trust, Chelmsford, Chelmsford, UK. <sup>46</sup>Royal Victoria Infirmary, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK. <sup>47</sup>Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK. <sup>48</sup>University of Aberdeen, Department of Applied Medicine, Aberdeen, Aberdeen, UK. <sup>49</sup>University Hospital Lewisham, Lewisham and Greenwich NHS Trust, Lewisham, UK. <sup>50</sup>Yeovil district hospital and Musgrove Park Hospital, Taunton and Somerset NHS Trust, Taunton and Somerset, UK. <sup>51</sup>University Hospital of North Durham, County Durham and Darlington NHS Foundation Trust, Durham, UK. <sup>52</sup>Glan Clwyd Hospital, Betsi Cadwaladr University Health Board, North Wales, UK.