

ARTICLE

Received 7 Jul 2016 | Accepted 27 Mar 2017 | Published 24 May 2017

DOI: 10.1038/ncomms15382

OPEN

Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants

Lam C. Tsoi^{1,2,3,*}, Philip E. Stuart^{1,*}, Chao Tian⁴, Johann E. Gudjonsson¹, Sayantan Das², Matthew Zawistowski², Eva Ellinghaus⁵, Jonathan N. Barker⁶, Vinod Chandran^{7,8}, Nick Dand⁶, Kristina Callis Duffin⁹, Charlotta Enerbäck¹⁰, Tõnu Esko^{11,12}, Andre Franke⁵, Dafna D. Gladman^{7,8}, Per Hoffmann^{13,14}, Külli Kingo¹⁵, Sulev Kõks^{16,17}, Gerald G. Krueger⁹, Henry W. Lim¹⁸, Andres Metspalu¹¹, Ulrich Mrowietz¹⁹, Sören Mucha⁵, Proton Rahman²⁰, Andre Reis²¹, Trilokraj Tejasvi^{1,22}, Richard Trembath²³, John J. Voorhees¹, Stephan Weidinger¹⁹, Michael Weichenthal¹⁹, Xiaoquan Wen², Nicholas Eriksson⁴, Hyun M. Kang², David A. Hinds⁴, Rajan P. Nair¹, Gonçalo R. Abecasis² & James T. Elder^{1,22}

Psoriasis is a complex disease of skin with a prevalence of about 2%. We conducted the largest meta-analysis of genome-wide association studies (GWAS) for psoriasis to date, including data from eight different Caucasian cohorts, with a combined effective sample size >39,000 individuals. We identified 16 additional psoriasis susceptibility loci achieving genome-wide significance, increasing the number of identified loci to 63 for European-origin individuals. Functional analysis highlighted the roles of interferon signalling and the NFκB cascade, and we showed that the psoriasis signals are enriched in regulatory elements from different T cells (CD8⁺ T-cells and CD4⁺ T-cells including T_{H0}, T_{H1} and T_{H17}). The identified loci explain ~28% of the genetic heritability and generate a discriminatory genetic risk score (AUC = 0.76 in our sample) that is significantly correlated with age at onset ($p = 2 \times 10^{-89}$). This study provides a comprehensive layout for the genetic architecture of common variants for psoriasis.

¹Department of Dermatology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ²Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. ³Department of Computational Medicine & Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ⁴23andMe, Inc., Mountain View, California 94041, USA. ⁵Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel 24105, Germany. ⁶St John's Institute of Dermatology, Division of Genetics and Molecular Medicine, Faculty of Life Sciences and Medicine, King's College London, London SE1 9RT, UK. ⁷Department of Medicine, Division of Rheumatology, University of Toronto, Toronto, Ontario, Canada M5S 1A8. ⁸Centre for Prognosis Studies in the Rheumatic Diseases, Toronto Western Research Institute, University of Toronto, Toronto, Ontario, Canada M5T 2S8. ⁹Department of Dermatology, University of Utah, Salt Lake City, Utah 84132, USA. ¹⁰Department of Dermatology, Linköping University, Linköping SE-581 83, Sweden. ¹¹Estonian Genome Center, University of Tartu, Tartu 51010, Estonia. ¹²Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ¹³Institute of Human Genetics, University of Bonn, Bonn 53127, Germany. ¹⁴Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel 4031, Switzerland. ¹⁵Dermatology Clinic, Tartu University Hospital, Department of Dermatology and Venereology, University of Tartu, Tartu 50417, Estonia. ¹⁶Department of Pathophysiology, Centre of Translational Medicine and Centre for Translational Genomics, University of Tartu, Tartu 50411, Estonia. ¹⁷Department of Reproductive Biology, Estonian University of Life Sciences, Tartu 51006, Estonia. ¹⁸Department of Dermatology, Henry Ford Hospital, Detroit, Michigan 48202, USA. ¹⁹Department of Dermatology, University Medical Center Schleswig-Holstein, Campus Kiel, Kiel 24105, Germany. ²⁰Memorial University, St. John's, Newfoundland and Labrador, Canada A1B 3X9. ²¹Institute of Human Genetics, FAU Erlangen-Nürnberg, Erlangen 91054, Germany. ²²Ann Arbor Veterans Affairs Hospital, Ann Arbor, Michigan 48105, USA. ²³Department of Medical and Molecular Genetics, King's College London, London WC2R 2LS, UK. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.T.E. (email: jelder@umich.edu).

Psoriasis is a chronic and complex multi-genic immune-mediated skin disease¹ affecting around 2% of European-origin individuals². Previous association studies of psoriasis have identified over 60 psoriasis susceptibility loci^{3–18}, 47 of which are associated with the risk of psoriasis in European-origin populations. These findings have greatly advanced the understanding of disease mechanisms and associated pathways. Thus, the *IL23R*, *IL12B*, *IL23A* and *TRAF3IP2* loci suggest a prominent role of the IL23 signalling pathway and promotion of T_H17 responses; whereas the *TNFAIP3*, *NFKBIA*, *NFKBIZ*, *TNIP1* and *RELA* loci suggest dysregulation of the NFκB pathway in disease pathogenesis^{19,20}. Approximately half (22) of the 47 European-origin loci were identified in cohorts containing a large proportion (≥50%) of samples genotyped using the Immunochip^{9,16,17,18}, a platform that focuses on genetic variants from promising signals identified in previous association studies of autoimmune diseases¹⁸. However, restricting analysis to markers genotyped (~110,000)⁹ or well-imputed (~700,000)¹⁷ on the Immunochip limits exploration of the full genome for susceptibility loci. Here, we present the largest genome-wide association study (GWAS) meta-analysis for psoriasis to date in European-origin individuals. We identified 16 new disease susceptibility regions, and we also revealed various functional networks and gene regulatory signals associated with psoriasis, providing novel insights into the immunopathogenesis of psoriasis.

Results

Meta-analysis of GWAS studies. We gathered genotype data from both published^{5–9,17,18} and new cohorts, consisting of seven genome-wide association studies (GWAS) and one Immunochip data set (Supplementary Table 1). The effective sample size of the combined data set (that is, the size of a sample with equal numbers of cases and controls that possesses equivalent statistical power to the meta-analysis) is ~40,000 samples (Supplementary Table 1); with >30,000 individuals for the GWAS component, it is about three times larger than the previous meta-analysis with the biggest GWAS component^{3,9,17}. After quality control, we performed genotype phasing²¹ and imputation²² using haplotypes from the 1,000 Genomes project²³. We then carried out logistic regression on each data set to determine genetic associations. Genomic inflation factors were below 1.05 for all data sets (Supplementary Table 1). Because the case definition for one of the GWAS cohorts (23andMe) was based on self-reported information, we used the risk allele frequencies for known loci in cases and controls in other cohorts to estimate the proportion of misclassified phenotypes (Supplementary Figs 1 and 2). Surprisingly, our results indicated that around 4% of the unaffected controls in the 23andMe cohort reported that they had psoriasis. To address this issue, we implemented a statistical approach²⁴ to adjust the summary statistics in this dataset for bias caused by response misclassification in logistic regression²⁵. As shown in Supplementary Fig. 3, this adjustment could correct for the downward bias of the estimated ORs and s.e.'s, at the cost of a substantial decrease in effective sample size.

We performed meta-analysis of 9,113,515 markers with good imputation quality²² ($r^2 \geq 0.7$) in at least four data sets, using the inverse-variance approach²⁶. Of 47 known psoriasis susceptibility loci, 42 (89%) achieved genome-wide significance ($p \leq 5 \times 10^{-8}$) (the remaining five yielded $p \leq 2 \times 10^{-4}$; Supplementary Table 2). Notably, we identified 16 new psoriasis susceptibility loci achieving genome-wide significance (Table 1 and Fig. 1; Supplementary Note; Supplementary Figs 4–6). Meta-analysis using all but the 23andMe dataset showed suggestive evidence ($p \leq 2 \times 10^{-5}$) for all new loci. Moreover, inclusion of the

23andMe data led to genome-wide significant findings for 14 of the 16 new loci, as only two loci achieved significance without the 23andMe data (Supplementary Table 3). Among the novel loci, the 19q13.33 region has been reported to reach genome-wide significance in a joint analysis of GWAS for Crohn's disease and psoriasis¹⁵, and in a contemporary analysis of exome array data. The adjusted ORs from the 23andMe cohort did not differ significantly from those of the other seven data sets ($p = 0.2$). We also found no significant heterogeneity of ORs among studies for all new loci (Cochran's Q p values >0.05). Furthermore, all new loci reached genome-wide significance under a random effects model²⁷.

Genetic architecture and risk scores. By increasing the number of European-origin psoriasis susceptibility loci to 63, we were able to explore the genetic architecture of psoriasis in greater detail. Interestingly, seven (44%) of the new loci were identified using only GWAS data sets (Supplementary Table 3), as the Immunochip data does not provide good genotype coverage of these regions. Moreover, only two new loci were identified in the 186 non-contiguous regions that underwent dense genotyping in the Immunochip platform²⁸. As shown for other complex traits²⁹, we found that the minor allele frequencies (MAFs) of the associated signals are negatively correlated ($\rho = -0.57$; $p = 2 \times 10^{-8}$) with the risk allele effect sizes of the disease loci (Fig. 2a; Supplementary Table 4; Supplementary Fig. 7). Thus, rs76959677 has the largest effect size (OR = 1.28) and the smallest MAF (=0.04) among the new loci (Table 1). Altogether, the 63 loci account for over 28% of the estimated heritability^{30,31}, as compared to 26% using only known loci. Our estimations are very similar to those obtained using other approaches, as shown in Supplementary Table 5. To evaluate whether the susceptibility loci could be used to discriminate between affected and unaffected individuals in our sample, we used the effect sizes and imputed dosages from our cohorts to compute genetic risk scores (GRS), and associated them with the disease status. Figure 2b shows receiver operating curves (ROC) plotting the true positive rate versus false positive rate under different GRS thresholds. The area under the curve (AUC) is 0.76, suggesting GRS has discriminative power for predicting disease status among individuals in these cohorts^{5–9,15}. Age-at-onset has emerged as a key clinical and stratification feature for psoriasis^{32–34}. To examine the correlation between age-at-onset and the GRS, we analysed 6,251 psoriatic patients for whom this information was available. Our results show that the GRS is inversely correlated with age-at-onset (Spearman $\rho = -0.25$; $p = 2 \times 10^{-89}$); mean age-at-onset was 34.9 years for psoriatic patients in the lowest fifth percentile of GRS, compared to 20.4 in those in the highest fifth percentile (Fig. 2c). This correlation remains significant after removing the MHC signal from the calculation ($\rho = -0.08$; $p = 2 \times 10^{-11}$).

Functional interpretation of GWAS data. To evaluate the underlying disease mechanisms responsible for these genetic signals, we applied a recently-developed algorithm termed minimum distance-based enrichment analysis for genetic association (MEAGA)³⁵ to simultaneously query biological functions and pathways, as well as protein-protein interactions, for enrichment among genes mapping to the identified psoriasis loci. We found 87 significantly enriched functions/pathways (false discovery rate ≤ 0.1 , Supplementary Table 6). As expected, many of these are immune-related functions such as lymphocyte differentiation/regulation, Type I interferon, pattern recognition and response to virus/bacteria (Fig. 3a; Supplementary Fig. 8). Among the enriched functions, 'Regulation of I-κB kinase/NF-κB cascade'

Table 1 | Newly identified psoriasis associated loci.

Chr	Pos	Marker	RA	NRA	RAF _{case}	RAF _{cont}	ORs	P value	Direction*	Nearby genes
1	78450517	rs34517439	A	C	0.13	0.12	1.18	4.43×10^{-9}	+ ? + + + + ? +	<i>FUBP1</i>
1	172675097	rs12118303	C	T	0.19	0.17	1.12	3.02×10^{-10}	+ + + + + + + +	<i>FASLG</i>
1	206655331	rs41298997	T	C	0.19	0.18	1.13	2.37×10^{-8}	+ + + + + + + +	<i>IKBKE</i>
10	64369999	rs2944542	G	C	0.62	0.60	1.08	1.76×10^{-8}	+ + + + - + + +	<i>ZNF365</i>
10	89824771	rs76959677	G	A	0.05	0.04	1.28	2.75×10^{-8}	+ + + + + ? +	<i>PTEN, KLLN, SNORD74</i>
10	102038641	rs61871342	G	A	0.57	0.55	1.10	1.56×10^{-9}	+ + - + + + ? +	<i>CHUK</i>
11	65593444	rs118086960	T	A	0.49	0.47	1.12	6.89×10^{-9}	+ + + + + ? +	<i>CFL1, FIBP, FOSL1</i>
12	10597207	rs11053802	T	C	0.69	0.67	1.11	4.17×10^{-9}	+ + + + + ? +	<i>KLRK1, KLRC4</i>
12	112059557	rs11065979	T	C	0.47	0.45	1.08	1.67×10^{-8}	+ + - + + + + +	<i>BRAP, MAPKAPK5</i>
12	122668326	rs11059675	A	G	0.48	0.46	1.10	1.50×10^{-8}	+ + + - + + ? +	<i>IL31</i>
13	99950260	rs9513593	G	A	0.19	0.18	1.12	3.60×10^{-8}	+ + + - + + + +	<i>UBAC2, RN7SKP9</i>
14	98668778	rs142903734	AAG	A	0.81	0.79	1.12	7.15×10^{-9}	+ + + + + + + +	<i>RP11-6101.1</i>
15	31637666	rs28624578	T	C	0.85	0.83	1.18	9.22×10^{-10}	+ + + + + ? +	<i>KLF13</i>
17	73890363	rs55823223	A	G	0.15	0.13	1.15	1.06×10^{-8}	+ + + + + + + +	<i>TRIM47, TRIM65</i>
18	12857002	rs559406	G	T	0.47	0.45	1.10	1.19×10^{-10}	+ + - + + + + +	<i>PTPN2</i>
19	49206417	rs492602	G	A	0.49	0.46	1.11	6.57×10^{-13}	+ + + + + + + +	<i>FUT2</i>

*Direction of the effect of the risk allele in the eight data sets in the order of: PsA GWAS, CASP GWAS, Kiel GWAS, Genizon GWAS, WTCCC2, Exomechip w/GWAS content, PAGE Immunochip and 23andMe GWAS, proceeding from left to right. '?' means the marker is not imputed well in the corresponding cohort. NRA, Non-risk allele; OR, odds ratio; RA, risk allele; RAF, risk allele frequency.

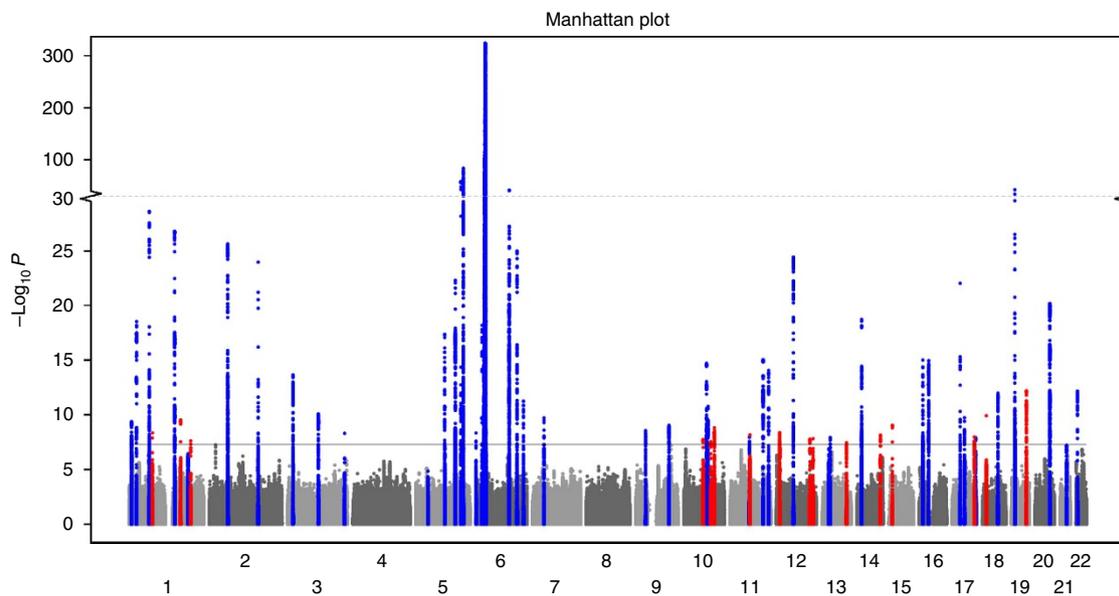


Figure 1 | Meta-analysis results. The ‘Manhattan’ plot shows the negative log *p* values of the meta-analysis. The known loci are coloured in blue; the sixteen novel loci are in red.

(Supplementary Fig. 9) contains genes from 11 associated loci, and three of them are novel (*CHUK* in 10q24.31, *IKBKE* in 1q32.1 and *FASLG* in 1q24.3). When taken together with genes (that is, *TNIP1*, *TNFAIP3* and *NFKBIA*) mapping to known psoriasis loci that are well-appreciated as components of NF- κ B signalling, our results further implicate this pathway in the pathophysiology of psoriasis. Other novel loci containing genes involved in the enriched functions identified by MEAGA include *KLRK1* (12p13.2) (‘regulation of leukocyte mediated cytotoxicity’) and *PTEN* (10q23.31) (‘regulation of response to external stimulus’). We next asked whether the observed association signals are enriched among gene regulatory regions that have been mapped to different immune cell subsets in publicly-available databases³⁶. Our results (Fig. 3b; Supplementary Table 7) show that the psoriasis signals are most enriched among enhancers in CD4⁺ T-helper (T_H0, T_H1 and T_H17) and CD8⁺ cytotoxic T cells, in concordance with the previous study³⁶. Indeed, thirteen of our novel loci (81%) either themselves harbour or are in high linkage disequilibrium ($r^2 \geq 0.8$) with SNPs mapping to enhancers in

these cell types (Supplementary Table 8). These results complement previous studies in providing functional characterization for psoriasis-associated loci^{37,38}. We then screened for existing drugs targeting genes from the psoriasis susceptibility loci in different drug databases^{39,40}. We found that seven genes from six novel loci are targets for 18 different drugs (Supplementary Table 9). Interestingly, some of these drugs (that is, aminosalicyclic acid⁴¹, mesalazine⁴² and sulfasalazine⁴³) have been used to treat psoriasis in clinical practice.

Discussion

Rather than relying on following up promising signals, here we show that the utilization of newer, less costly GWAS assays to interrogate the entire genome in follow-up samples is a cost-effective approach capable of revealing subtle genetic signals. In addition, we have implemented an approach used in epidemiology studies to adjust a misclassified binary outcome²⁴. To our knowledge, this is the first large genetic association study to compare outcomes using specialist-diagnosed versus self-

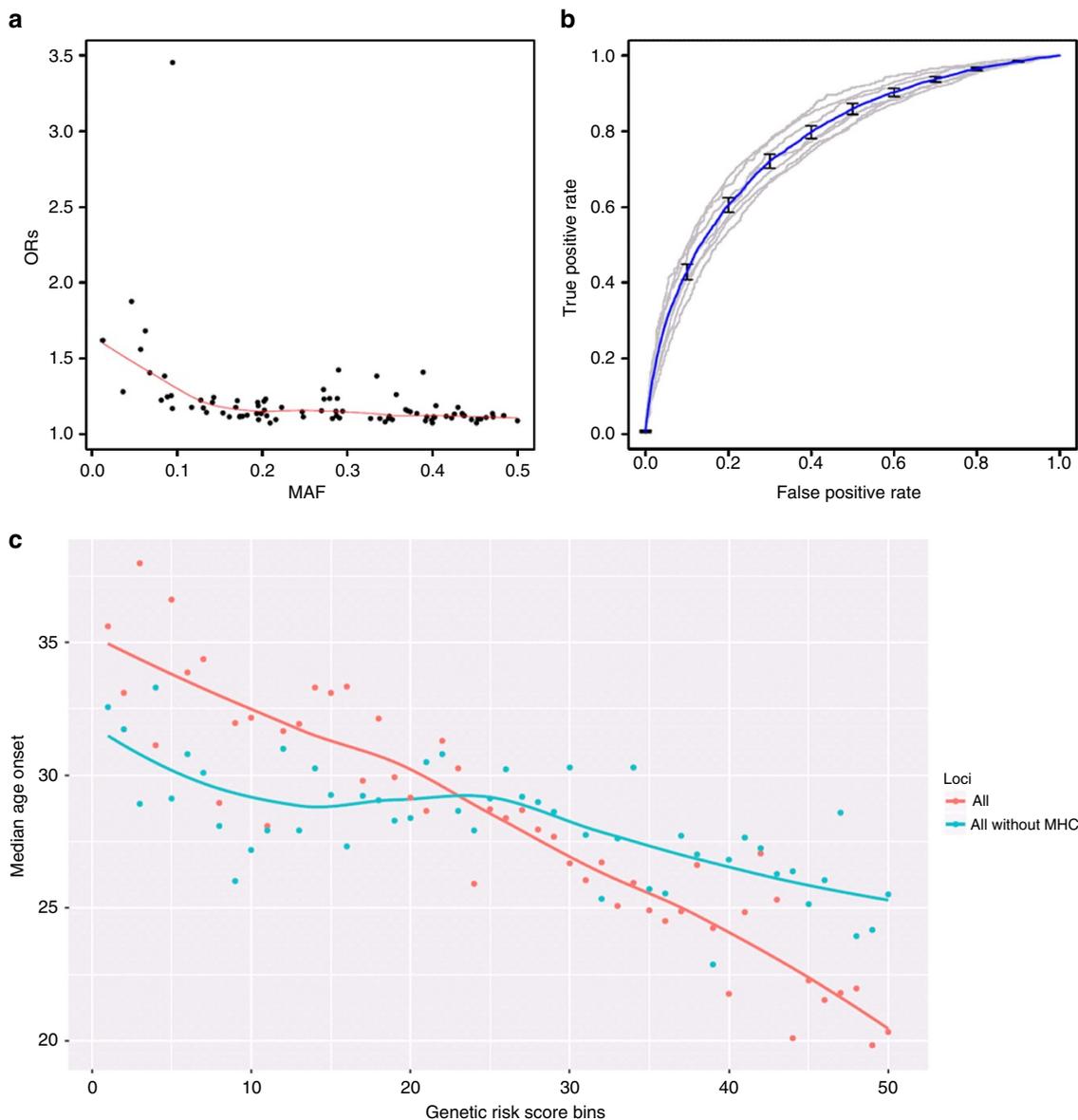


Figure 2 | Association of psoriasis susceptibility with disease risk. (a) The effect size (odds ratio, OR) of the risk allele is plotted against the minor allele frequency of the signal among all susceptibility loci. (b) True positive rate versus false positive rate for using genetic risk score to distinguish psoriasis versus control samples. Blue line shows the averaged results among the different cohorts (grey), and the s.e. bars are also shown. (c) The median age-at-onset of psoriasis is plotted against different percentile bins (every 2%) of genetic risk scores for all loci (blue) or all loci without MHC (red).

reported affection and to adjust for response misclassification. Of note, we observed that disease allele frequencies and ORs were underestimated in an independent study that defined psoriasis status based on the electronic health records⁴⁴. This may be because psoriatic lesions appear similar to other common skin diseases, including atopic eczema and seborrhoeic dermatitis, leading to misdiagnosis. Our results illustrate the importance of correcting misclassification of disease outcome as large-scale data-mining of phenotypes becomes more common. The disease-associated loci define a GRS that is capable of discriminating case-control status in our sample (AUC = 0.76). Similar results have been reported in the Chinese population⁴⁵ as well as a smaller European-origin sample⁴⁶. In concordance with previous studies^{45,46}, we found that the GRS is also strongly inversely correlated with age-at-onset of psoriasis, with the MHC comprising much of this effect (Fig. 2c). The strong association between HLA-Cw6 and streptococcal infection in juvenile-onset

psoriasis may explain part of this association⁴⁷. However, correlations between genetic risk allele load and age-at-onset are not universal in complex genetic disorders^{48,49}, and the relationship between GRS and age-at-onset needs to be explored on a disease-by-disease basis. While we did not find any disease-associated variants that alter protein structure in new loci, we demonstrated significant enrichment for genes involved in immune system function among the known and novel genetic signals. We also found significant enrichment of psoriasis genetic signals in active chromatin domains in Th1 and Th17 cells (Fig. 3). Among the individual candidates (Supplementary Note), *FASLG* encoding Fas ligand, *IKBKE* encoding IKK- ϵ , *CHUK* encoding IKK- α , *IL31* encoding the cytokine IL-31, *KLRK1* encoding NKG2D, a killer cell lectin-like receptor and *PTPN2* encoding T-cell protein tyrosine phosphatase, all play prominent roles in T-cell activation, signalling and/or effector function. By guiding further functional investigation into the roles of these

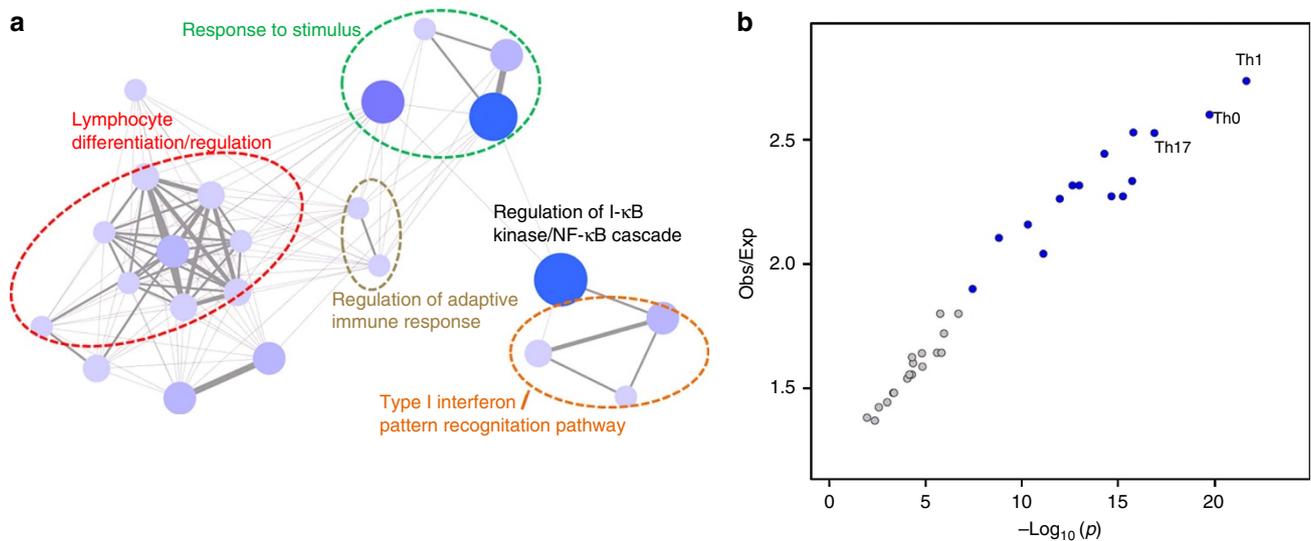


Figure 3 | Biological inference for the psoriasis loci. (a) Enriched functions (nodes) among the associated loci identified using MEAGA. For illustration purposes, only functions with at least four genes overlapping with other enriched functions are shown (the full list of enriched functions/pathways is shown in Supplementary Table 6). The size of the nodes and the width of the edges correlate with the number of overlapped disease-loci and the number of shared disease-loci, respectively. Nodes with dark blue colour represents higher numbers of overlapped loci while lighter colour represents lower numbers of overlapped loci. The functional annotations for the nodes are presented in Supplementary Fig. 8. (b) The observed-to-expected ratio of the number of regulatory-element overlapped loci versus the enrichment p value. Immune cells are highlighted in blue.

variants in the regulation of their target genes, as well as further functional investigation of these targets, these results will serve as an important framework guiding future research into the pathogenesis and treatment of psoriasis.

Methods

Data sets. The collection of samples for the five GWAS (PsA, CASP, Kiel, Genizon and WTCCC2) and the Immunochip dataset were described previously^{5–7,9,18}. The new Exomechip cohort, consisting of 6,463 genetically independent psoriasis cases and 6,096 unrelated controls of European Caucasian descent collected in North America and Sweden, was genotyped using the Affymetrix Axiom Biobank Plus Genotyping Array at the Affymetrix facility (Santa Clara, CA). All human subjects provided written informed consent and were enrolled according to the protocols approved by the institutional review board for human subject research of each institution, in adherence with the Declaration of Helsinki principles. The basic array contains 246,000 genome-wide markers, 265,000 exome coding SNPs and indels, and 95,000 eQTL, pharmacogenomic and novel loss-of-function variants, which was supplemented by addition of 77,000 custom markers consisting of (1) 50,000 novel rare variants identified in targeted sequencing of known psoriasis loci, (2) 15,000 additional evenly distributed markers to enhance GWAS coverage, (3) 10,000 1,000 Genome variants in ENCODE-predicted regulatory regions for normal human epidermal keratinocyte (NHK) cell lines, (4) 1,000 markers associated with other autoimmune diseases and (5) 1,000 markers representing skin eQTLs identified from our analyses of the psoriatic skin transcriptome. Psoriasis cases in these cohorts were dermatologist-diagnosed, and all studies were approved by the ethical committees of their respective institutions. In this study, we did not segregate psoriasis subphenotypes (that is, psoriatic arthritis, cutaneous-only psoriasis¹⁸); therefore our cohorts include psoriasis cases that might have developed psoriatic arthritis, and this is especially true for the PsA GWAS, in which all included cases have psoriatic arthritis.

Quality control. For each dataset, we removed samples with high missingness ($>2\%$) or a high inbreeding coefficient ($|F| > 0.03$), and we also removed markers with low call rate ($<95\%$), with more than two alleles, or that failed Hardy Weinberg equilibrium ($p < 1 \times 10^{-6}$). We identified duplicated or highly related pairs (that is, first and second degree relatives) of individuals among our data sets using independent markers outside of the known psoriasis susceptibility loci ('null markers'⁹); this includes samples that were genotyped in multiple cohorts (for example, the same sample might be genotyped in both the CASP GWAS or Exomechip cohorts). When related or identical pairs were identified in different data sets, we preferentially kept the sample from the genotyping platform with the higher number of markers with genome-wide coverage (Supplementary Table 1). We used the independent (that is, $ld-r^2 < 0.2$) markers that are outside the known psoriasis loci to compute the principal components for each data set; and for the Immunochip data set, since the platform is enriched with markers

from the immune-associated regions, we first conducted a meta-analysis using the CASP, Kiel, and WTCCC2 cohorts and identified independent markers which have meta-analysis P values > 0.5 as 'null markers' to compute the principal components. We then used principal components to remove the population outliers to ensure all analysed individuals were of European ancestry⁹.

23andMe cohort. The 23andMe cohort was drawn from the customer base of 23andMe, Inc., a personal genetics company. The samples from this cohort were genotyped on one of four platforms: the V1 and V2 platforms were variants of the Illumina HumanHap550 BeadChip with additional custom content; the V3 platform is a variant of the Illumina OmniExpress + BeadChip, with custom content; the V4 platform is a fully custom design, including lower redundancy subsets of V2 and V3 SNPs with coverage of low allele frequency coding variants, as well as 570,000 additional SNPs. Research participants included in the cohort provided informed consent and answered surveys online according to the 23andMe human subject protocol, which was reviewed and approved by Ethical & Independent Review Services, a private institutional review board. The 'psoriasis' phenotype combines self-reported psoriasis diagnoses from several sources available on the 23andMe website: (i) Medical History Survey; (ii) Roots into the future intake form; (iii) research snippet. There are three choices (yes, no, not sure) for each psoriasis-related question from each source. We merged the yes/no responses for these questions, with inconsistent responses scored as missing: cases have at least one positive response and no negative responses, and controls have at least one negative response and no positive responses. We also derived responses from two additional questions derived from the IBD Community Survey and Health Intake Form, regarding whether the individual has been diagnosed with psoriasis to define cases (when any response is a yes) and controls (when it is not a case and at least one response is control).

Imputation and association. We performed haplotype phasing²¹ and imputation²² for each dataset. For imputation, we used haplotypes from all populations in the 1,000 Genomes Project phase 1 (release 3) as a reference panel²³. We then analysed markers with imputation quality greater than 0.7 in at least half (that is, 4) of the data sets. For each data set, we performed logistic regression using top principal components and data collection center indicator variables as covariates to correct for population stratification. We computed the inflation factor (λ) using the 'null markers' for the genomic control analysis (Supplementary Table 1).

Proportion of true positives among 23andMe psoriasis cases. For each of the previously identified signals from the known psoriasis loci, we compared the risk allele frequencies in cases and controls estimated from our dermatologist diagnosed-based data⁹ with those estimated by the 23andMe cohort. The RAFs in cases from the dermatologist-diagnosed cohorts are systematically higher than those in the 23andMe cohort (34 out of 36 loci listed in Tsoi *et al.*⁹ manifested

RAF_{case_Tsoi(2012)} higher than those estimated in RAF_{case_23andMe}); while the RAFs in controls are highly concordant (Supplementary Fig. 1). We hypothesized that some of the defined cases are false positives (that is, the individuals do not actually have psoriasis). Assuming the defined cases from the 23andMe cohort contain a mixture of true cases and controls, we would get:

$$\text{RAF}_{\text{case23 and Me}} = (q)\text{RAF}_{\text{caseTsoi (2012)}} + (1 - q)\text{RAF}_{\text{controlTsoi (2012)}}$$

where q is the proportion of true positives. The proportion could then be estimated as:

$$\text{Median}_{\text{known loci}} \left(\frac{\text{RAF}_{\text{case23 and Me}} - \text{RAF}_{\text{controlTsoi (2012)}}}{\text{RAF}_{\text{caseTsoi (2012)}} - \text{RAF}_{\text{controlTsoi (2012)}}} \right)$$

We estimated that $q = 0.36$. Ignoring the misclassification of psoriasis phenotype, the 16,120 self-reported cases and 254,909 controls of the 23andMe cohort yield an estimated disease prevalence of 5.9%. But if we assume $q = 0.36$ and correct for misclassification, then we would obtain a 2.1% prevalence (Supplementary Table 1), matching the disease frequencies estimated for European-origin populations⁵⁰.

Adjustment for misclassification of 23andMe cases. We employed Duffy's approach to adjust odds ratios and s.e.'s for bias caused by response misclassification in logistic regression²⁴. If β^* and $V(\beta^*)$ are the naive log OR and its variance for the misclassified case-control data, then by Duffy's method the corrected log OR and its variance can be estimated as:

$$\beta = \beta^* / \left\{ \frac{(\alpha_1 - p^*)(\alpha_2 - 1 + p^*)}{(\alpha_1 + \alpha_2 - 1)p^*(1 - p^*)} \right\}$$

and

$$V(\beta) = V(\beta^*) \left\{ \frac{(\alpha_1 + \alpha_2 - 1)p^*(1 - p^*)}{(\alpha_1 - p^*)(\alpha_2 - 1 + p^*)} \right\}^2 + (\beta^* p^*(1 - p^*))^2 \left\{ \frac{V(\alpha_1)}{(\alpha_1 - p^*)^4} + \frac{V(\alpha_2)}{(\alpha_2 - 1 + p^*)^4} \right\}$$

Parameters α_1 and α_2 are the sensitivity and specificity of the binary classification. For the 23andMe sample we assumed $\alpha_1 = 1$ (that is, all true cases were reported as such); using $q = 0.36$, α_2 could then be estimated as 0.9611. $V(\alpha_2)$ was estimated to be 2.67×10^{-6} using Monte Carlo simulation based on the observed RAFs for 32,240 case chromosomes from the 23andMe cohort and for 21,176 case and 45,612 control chromosomes from our previous study⁹. $V(\alpha_1)$ was assumed to be 0. The observed case prevalence in the sample (p^*) is 0.0595. Because p^* is small, deviations of our assumptions for α_1 and $V(\alpha_1)$ from their true values have little impact on the resulting estimates of β and $V(\beta)$, which was verified by sensitivity analysis for a broad range of both parameters (that is, $\alpha_1 = 0.5 - 1.0$ and $V(\alpha_1) = 0 - 0.001$).

Meta-analysis and effective sample size calculation. We used the inverse-variance approach implemented in METAL²⁶ to perform meta-analysis across the eight data sets. The effective sample size for each dataset (except for the 23andMe study) was computed as: $N_{\text{eff}} = \frac{4}{\frac{1}{N_{\text{case}}} + \frac{1}{N_{\text{controls}}}}$. The effective sample size approach was computed for each cohort to provide an estimate of the sample size corresponding to an equal case/control balance, and has the same statistical power to identify true association. To compute the effective sample size for the 23andMe study, we first determined the asymptotic relative efficiency (ARE) of the Duffy-corrected log OR to the log OR that would have been obtained if there had been no misclassification of disease phenotype. Because both of these OR estimators are convergent and asymptotically unbiased, the ARE for these two parameters (and their corresponding Wald chi-square test statistics) is equal to the ratio of their variances. We determined this variance ratio by simulation. We bootstrap sampled one of our largest studies with dermatologist-diagnosed phenotypes (the PAGE Immunochip study) to create 25 data sets mimicking the 23andMe study; that is, each simulated dataset had 5,803 true cases, 10,317 false cases, and 254,909 true controls. For each of 28 independent known psoriasis loci with adequate significance of association in the PAGE Immunochip ($p < 0.001$), we determined the mean variance across the 25 data sets for the Duffy-corrected ORs from analysis of the misclassified responses and also the mean variance for the ORs from analysis of the true responses. The ARE for each locus was the ratio of these two mean variances, and the median ARE for the 28 loci was 0.35. The effective sample size for the 23andMe study after using Duffy's adjusted approach was then estimated by multiplying the effective sample size under no misclassification ($22,715 = 4/(1/5,803 + 1/265,226)$) by the ARE, yielding an estimate of 7,950 (Supplementary Table 1).

Analysis of psoriasis loci. We performed comparisons between the estimated risk allele ORs versus MAFs of the best associated markers for each of the 63 loci. In a recent fine-mapping meta-analysis study (manuscript in preparation), we observed that a substantial proportion of psoriasis loci harbor secondary independent signals, and due to the linkage disequilibrium structure the effect sizes from the primary signals could be over/under estimated if the risk allele of secondary

signals tend to be on the same haplotype of the risk/non-risk alleles of the primary signals, respectively. Therefore, for each published locus with multiple signals, we computed the ORs by conditioning on the other independent signal(s) (that is, as covariates) from the same locus. These values were also used in the calculation of the variance in liability explained³¹. We next computed the genetic risk score using the effect sizes and imputed dosage data for each signal. Let the OR of the risk allele in signal i be OR_i , and the imputed dosage/genotypes for the risk allele be d_i , the genetic risk score for an individual among all independent associated signals is computed as:

$$\sum_{i=1}^L d_i \ln(OR_i)$$

Enrichment analysis. We analysed our GW-significant loci using MEAGA³⁵. MEAGA employs a graphical algorithm to measure the closeness between gene-set overlapping genes in the biological interactome, to identify the enriched functions/pathways among the 63 psoriasis loci. We used the protein-protein interaction data from BioGRID⁵¹ for generating the interactome, and the nine million markers examined in this meta-analysis as background under the default setting in MEAGA for the enrichment analysis. 50,000 samplings were used for P value estimation. We then sought to understand whether the psoriasis signals are enriched among regulatory elements in different cell types, as predicted by H3K27ac chromatin marks. We utilized the active enhancers identified from a recent study aiming to use epigenomics data to fine-map genetic susceptibility loci for complex autoimmune diseases³⁶. The 33 cell types under study included: T_{naive} , T_{mem} , T_{reg} , T_{stim} , Th17, Th1, Th0, Th2, CD8_{naive}, CD8_{mem}, Monocytes, B cell, Lymphoblastoid, B centroblast, CD34⁺, K562, Inferior temporal lobe, Angular gyrus, Mid-frontal lobe, Cingulate gyrus, Substantia nigra, Anterior caudate, Hippocampus middle, Colonic mucosa, Duodenum mucosa, Adipose, HepG2, Liver, Pancreatic islets, Kidney, Human skeletal muscle myoblasts (HSMM), NH osteoblast, Chondrogenic diff. We performed enrichment analysis by first enumerating the number of associated loci that overlap or are in linkage disequilibrium (LD) ($r^2 \geq 0.8$) with markers in regulatory elements, and then comparing that with the expected number of overlaps. The expected numbers were estimated by randomly sampling markers from the meta-analysis matching the LD-block length, MAF, and the number of genes in the LD-block, and counting the number of times these null markers overlap/in LD with the regulatory elements.

Drug databases. We downloaded data from PharmGKB⁴⁰ and Drugbank³⁹, and searched for drugs with potential gene targets from these databases.

Data availability. The data of the Exomechip cohort is available in dbGap (phs001306.v1.p1). The GWAS statistics from the 23andMe cohort can be requested by applying to the 23andMe collaboration program.

References

- Lowes, M. A., Suarez-Farinas, M. & Krueger, J. G. Immunology of psoriasis. *Annu. Rev. Immunol.* **32**, 227–255 (2014).
- Chandran, V. & Raychaudhuri, S. P. Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis. *J. Autoimmun.* **34**, J314–J321 (2010).
- Yin, X. *et al.* Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. *Nat. Commun.* **6**, 6916 (2015).
- Zuo, X. *et al.* Whole-exome SNP array identifies 15 new susceptibility loci for psoriasis. *Nat. Commun.* **6**, 6793 (2015).
- Ellinghaus, E. *et al.* Genome-wide association study identifies a psoriasis susceptibility locus at TRAF3IP2. *Nat. Genet.* **42**, 991–995 (2010).
- 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).
- 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).
- Stuart, P. E. *et al.* Genome-wide association analysis identifies three psoriasis susceptibility loci. *Nat. Genet.* **42**, 1000–1004 (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).
- Sun, L. D. *et al.* Association analyses identify six new psoriasis susceptibility loci in the Chinese population. *Nat. Genet.* **42**, 1005–1009 (2010).
- Li, Y. *et al.* Association analyses identifying two common susceptibility loci shared by psoriasis and systemic lupus erythematosus in the Chinese Han population. *J. Med. Genet.* **50**, 812–818 (2013).
- Sheng, Y. *et al.* Sequencing-based approach identified three new susceptibility loci for psoriasis. *Nat. Commun.* **5**, 4331 (2014).
- Tang, H. *et al.* A large-scale screen for coding variants predisposing to psoriasis. *Nat. Genet.* **46**, 45–50 (2014).
- 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).

15. 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).
16. Bowes, J. *et al.* Dense genotyping of immune-related susceptibility loci reveals new insights into the genetics of psoriatic arthritis. *Nat. Commun.* **6**, 6046 (2015).
17. Tsoi, L. C. *et al.* Enhanced meta-analysis and replication studies identify five new psoriasis susceptibility loci. *Nat. Commun.* **6**, 7001 (2015).
18. Stuart, P. E. *et al.* Genome-wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture. *Am. J. Hum. Genet.* **97**, 816–836 (2015).
19. Gudjonsson, J. E. & Elder, J. T. in *Dermatology in General Medicine*, Vol. 1 (ed. Goldsmith, L. *et al.*), 197–231 (McGraw-Hill, 2012).
20. Harden, J. L., Krueger, J. G. & Bowcock, A. M. The immunogenetics of Psoriasis: a comprehensive review. *J. Autoimmun.* **64**, 66–73 (2015).
21. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
22. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
23. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
24. Duffy, S. W. *et al.* A simple model for potential use with a misclassified binary outcome in epidemiology. *J. Epidemiol. Community Health* **58**, 712–717 (2004).
25. Albert, P. S., Liu, A. & Nansel, T. Efficient logistic regression designs under an imperfect population identifier. *Biometrics* **70**, 175–184 (2014).
26. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
27. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
28. Cortes, A. & Brown, M. A. Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.* **13**, 101 (2011).
29. Park, J. H. *et al.* Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc. Natl Acad. Sci. USA* **108**, 18026–18031 (2011).
30. Grijbovski, A. M., Olsen, A. O., Magnus, P. & Harris, J. R. Psoriasis in Norwegian twins: contribution of genetic and environmental effects. *J. Eur. Acad. Dermatol. Venereol.* **21**, 1337–1343 (2007).
31. So, H. C., Gui, A. H., Cherny, S. S. & Sham, P. C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.* **35**, 310–317 (2011).
32. Gladman, D. D., Antoni, C., Mease, P., Clegg, D. O. & Nash, P. Psoriatic arthritis: epidemiology, clinical features, course, and outcome. *Ann. Rheum. Dis.* **64**, iii14–iii17 (2005).
33. Henseler, T. & Christophers, E. Psoriasis of early and late onset: characterization of two types of psoriasis vulgaris. *J. Am. Acad. Dermatol.* **13**, 450–456 (1985).
34. Queiro, R., Tejon, P., Alonso, S. & Coto, P. Age at disease onset: a key factor for understanding psoriatic disease. *Rheumatology* **53**, 1178–1185 (2014).
35. Tsoi, L. C., Elder, J. T. & Abecasis, G. R. Graphical algorithm for integration of genetic and biological data: proof of principle using psoriasis as a model. *Bioinformatics* **31**, 1243–1249 (2015).
36. Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
37. Lin, Y. *et al.* Identification of cell types, tissues and pathways affected by risk loci in psoriasis. *Mol. Genet. Genomics* **291**, 1005–1012 (2016).
38. Aterido, A. *et al.* Genome-wide pathway analysis identifies genetic pathways associated with Psoriasis. *J. Invest. Dermatol.* **136**, 593–602 (2016).
39. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–D1097 (2014).
40. Whirl-Carrillo, M. *et al.* Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* **92**, 414–417 (2012).
41. Cortes-Franco, R., Fernandez, L. M. & Dominguez-Soto, L. Psoriasis treatment with 5-aminosalicylic acid. *Int. J. Dermatol.* **33**, 573–575 (1994).
42. Bharti, R. Mesalazine in treatment of psoriasis. *Indian J. Dermatol. Venereol. Leprol.* **62**, 231–232 (1996).
43. Gupta, A. K. *et al.* Sulfasalazine improves psoriasis. A double-blind analysis. *Arch. Dermatol.* **126**, 487–493 (1990).
44. Denny, J. C. *et al.* Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1110 (2013).
45. Yin, X. *et al.* A weighted polygenic risk score using 14 known susceptibility variants to estimate risk and age onset of psoriasis in Han Chinese. *PLoS ONE* **10**, e0125369 (2015).
46. Chen, H. *et al.* A genetic risk score combining ten psoriasis risk loci improves disease prediction. *PLoS ONE* **6**, e19454 (2011).
47. Gudjonsson, J. E. *et al.* Distinct clinical differences between HLA-Cw*0602 positive and negative psoriasis patients—an analysis of 1019 HLA-C- and HLA-B-typed patients. *J. Invest. Dermatol.* **126**, 740–745 (2006).
48. Scherr, R., Essers, J., Hakonarson, H. & Kugathasan, S. Genetic determinants of pediatric inflammatory bowel disease: is age of onset genetically determined? *Dig. Dis.* **27**, 236–239 (2009).
49. Sorosina, M. *et al.* Inverse correlation of genetic risk score with age at onset in bout-onset and progressive-onset multiple sclerosis. *Mult. Scler.* **21**, 1463–1467 (2014).
50. Parisi, R., Symmons, D. P., Griffiths, C. E. & Ashcroft, D. M. Global epidemiology of psoriasis: a systematic review of incidence and prevalence. *J. Invest. Dermatol.* **133**, 377–385 (2012).
51. Chahr-Aryamontri, A. *et al.* The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* **43**, D470–D478 (2015).

Acknowledgements

This work was supported by awards from the National Institutes of Health (R01AR042742, R01AR050511, R01AR054966, R01AR063611, R01AR065183 to J.T.E., as well as a GAIN award from the Foundation for the National Institutes of Health to GRA). L.C.T. is supported by awards from the Dermatology Foundation, the National Psoriasis Foundation and the Arthritis National Research Foundation. J.T.E. and T.T. are supported by the Ann Arbor Veterans Affairs Hospital, and L.C.T., P.E.S., T.T., J.E.G., J.J.V., R.P.N. and J.T.E. are supported by the Dawn and Dudley Holmes Foundation and the Babcock Memorial Trust. R.C.T. and J.N.B. were supported by the Medical Research Council Stratified Medicine award (MR/L011808/1). A.F. and E.E. received infrastructure support through the DFG Clusters of Excellence 306 'Inflammation at Interfaces' and is supported by the German Ministry of Education and Research (BMBF) through the e:Med sysINFLAME grant. J.E.G. was supported by Doris Duke Foundation (Grant #:2013106) and the National Institute of Health (K08AR060802 and R01AR06907) and the Taubman Medical Research Institute as the Frances and Kenneth Eisenberg Emerging Scholar. We acknowledge support from the Department of Health via the NIHR comprehensive Biomedical Research Center award to GSTT NHS Foundation Trust in partnership with King's College London and KCH NHS Foundation Trust. We acknowledge the Collaborative Association Study of Psoriasis (CASP) and the Wellcome Trust Case Control Consortium 2 (WTCCC2) for the contribution of GWAS data, as well as the provision of control DNA samples by the Cooperative Research in the Region of Augsburg (KORA) and genotyping data generated by the Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) consortium. We thank the Barbara and Neal Henschel Charitable Foundation for their support of the National Psoriasis Victor Henschel BioBank. The Heinz Nixdorf Recall (HNR) cohort was established with the support of the Heinz Nixdorf Foundation. The Estonian Psoriasis cohort was supported by institutional research funding IUT20-46 of the Estonian Ministry of Education and Research, by the Centre of Translational Genomics of University of Tartu (SP1GVARENG) and by the European Regional Development Fund (Centre of Translational Medicine, University of Tartu). The genotyping of control individuals from the HNR cohort was financed by the German Federal Ministry of Education and Research (BMBF), within the context of the NGFNplus Integrated Genome Research Network MooDS (Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia). We also thank the Psoriasis Association Genetics Extension (PAGE) for the contribution of Immunochip data. We would like to thank the research participants and employees of 23andMe for making this work possible. 23andMe's contribution to this work was supported by the National Human Genome Research Institute of the National Institutes of Health (grant number R44HG006981). We acknowledge the support of the International Psoriasis Council through a grant to J.N.B., G.R.A., J.T.E., A.F. and R.C.T. towards genotyping costs. The authors dedicate this article to the late Dr. Tilo Henseler, who played a pioneering role in developing our understanding of the genetics of psoriasis.

Author contributions

L.C.T., P.E.S. and J.T.E. designed the study. L.C.T. conducted the imputation, meta-analysis and downstream functional analysis. L.C.T., P.E.S., M.Z., X.W., H.M.K., G.R.A. worked on the bias adjustment method and statistical genetic aspects of the study. S.D. and L.C.T. conducted the exomechip quality control. J.E.G., E.E., J.N.B., V.C., N.D., K.C.D., C.E., T.E., A.F., D.D.G., P.H., K.K., S.K., G.G.K., H.W.L., A.M., U.M., S.M., P.R., A.R., T.T., R.T., J.J.V., S.W., M.W., R.P.N., J.T.E. contributed to the collection/data coordination of the samples in the cohorts. J.E.G. and J.T.E. provided biological inference of the candidate genes for each new locus. L.C.T. and J.T.E. wrote the first draft of the manuscript, and every author has reviewed the work.

Additional information

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

Competing interests: C.T. and D.A.H. are employees of and own stock options in 23andMe, Inc. N.K.E. was an employee of 23andMe when the study was conducted. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Tsoi, L. C. *et al.* Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. *Nat. Commun.* **8**, 15382 doi: 10.1038/ncomms15382 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017