Machine learning applied to atopic dermatitis transcriptome reveals distinct therapy-dependent modification of the keratinocyte immunophenotype.

K. Clayton, MRes\textsuperscript{1}, A. Vallejo, PhD\textsuperscript{1}, S. Sirvent, PhD\textsuperscript{1}, J. Davies, MRes\textsuperscript{1}, G. Porter, MSc\textsuperscript{1}, F. Lim, PhD\textsuperscript{4}, M.R. Ardem-Jones, FRCP\textsuperscript{1,3}*, M.E. Polak, PhD\textsuperscript{1,2}**

Current affiliations and addresses:
\textsuperscript{1} Clinical and Experimental Sciences, Sir Henry Wellcome Laboratories, Faculty of Medicine, University of Southampton, Southampton, United Kingdom
\textsuperscript{2} Institute for Life Sciences, University of Southampton, United Kingdom
\textsuperscript{3} Department of Dermatology, University Hospitals Southampton NHS Foundation Trust, United Kingdom
\textsuperscript{3} Unilever, Colworth Science Park, Sharnbrook, Bedford, United Kingdom
* Corresponding authors
MEP: m.e.polak@soton.ac.uk MAJ: m.aj@soton.ac.uk
** Contributed equally

Word count: 2993 Figure count: 4 Table count: 0

Funding declaration:
This work was supported by the Medical Research Council (MRC) DTP PhD Scheme and iCASE partnership with Unilever (scheme no. MR/N014308/1). MEP is funded by Sir Henry Dale Fellowship, Wellcome Trust (Grant no. 10937/Z/15/Z).
The authors declare they have no conflict of interest.

Conflict of interest:
None declared by any author
Abstract

Background
Atopic dermatitis (AD) arises from a complex interaction between an impaired epidermal barrier, environmental exposures, and the infiltration of Th1/Th2/Th17/Th22 T cells. Transcriptomic analysis has advanced understanding of gene expression in cells and tissues. However, molecular quantitation of cytokine transcripts does not predict the importance of a specific pathway in AD or cellular responses to different inflammatory stimuli.

Objective
To understand changes in keratinocyte transcriptomic programmes in human cutaneous disease during development of inflammation and in response to treatment.

Methods
We performed in silico deconvolution of the whole-skin transcriptome. Using co-expression clustering and machine learning tools, we resolved the gene expression of bulk skin (n=7 datasets, n=406 samples), firstly, into unsupervised keratinocyte immune response phenotypes and, secondly, into 19 cutaneous cell signatures of purified populations from publicly available datasets.

Results
We identify three unique transcriptomic programmes in keratinocytes, KC1, KC2, KC17, characteristic to immune signalling from disease-associated helper T cells. We cross-validate those signatures across different skin inflammatory conditions and
disease stages and demonstrate that the keratinocyte response during treatment is therapy dependent. Broad spectrum treatment with ciclosporin ameliorated the KC17 response in AD lesions to a non-lesional immunophenotype, without altering KC2. Conversely, the specific anti-Th2 therapy, dupilumab, reversed the KC2 immunophenotype.

**Conclusion**

Our analysis of transcriptomic signatures in cutaneous disease biopsies reveals the complexity of keratinocyte programming in skin inflammation and suggests that the perturbation of a single axis of immune signal alone may be insufficient to resolve keratinocyte immunophenotype abnormalities.
Abbreviations

AD Atopic dermatitis
AD-AcL Lesional (acute) AD skin
AD-ChL Lesional (chronic) AD skin
AD-Les Lesional (acute or chronic) AD skin
AD-Non Non-lesional AD skin
CD4 Cluster of differentiation 4
CD8 Cluster of differentiation 8
DDC CD11c⁺ Dermal Dendritic Cell
FDR False Discovery Rate
GCRMA Gene Chip Robust Multiarray Averaging
GO Gene Ontology
GOBP Gene Ontology, Biological Process
HH Healthy
IFN Interferon
IL-1 Interleukin-1
IL-13 Interleukin-13
IL-17 Interleukin-17
IL-22 Interleukin-22
IL-26 Interleukin-26
IL-4 Interleukin-4
IL4Rα Interleukin-4 Receptor, alpha subunit
KC Keratinocytes
KC1 Keratinocyte transcriptomic programme to type 1 inflammation
KC17 Keratinocyte transcriptomic programme to type 17 inflammation
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>KC2</td>
<td>Keratinocyte transcriptomic programme to type 2 inflammation</td>
</tr>
<tr>
<td>KC22</td>
<td>Keratinocyte transcriptomic programme to type 22 inflammation</td>
</tr>
<tr>
<td>KFG</td>
<td>Keratinocyte Growth Factor</td>
</tr>
<tr>
<td>KRT6A</td>
<td>Keratin 6A gene</td>
</tr>
<tr>
<td>KRT6B</td>
<td>Keratin 6B gene</td>
</tr>
<tr>
<td>LC</td>
<td>Langerhans Cell</td>
</tr>
<tr>
<td>LIMMA</td>
<td>Linear Models for Microarray Data</td>
</tr>
<tr>
<td>MCL</td>
<td>Markov Clustering algorithm</td>
</tr>
<tr>
<td>Ps-Les</td>
<td>Lesional psoriatic skin</td>
</tr>
<tr>
<td>Ps-Non</td>
<td>Non-lesional psoriatic skin</td>
</tr>
<tr>
<td>S100</td>
<td>S100 protein family gene</td>
</tr>
<tr>
<td>SCORAD</td>
<td>SCORing AD index</td>
</tr>
<tr>
<td>SCORAD50</td>
<td>50% improvement in baseline SCORAD index</td>
</tr>
<tr>
<td>SERPIN</td>
<td>Serine protease inhibitor family gene</td>
</tr>
<tr>
<td>Th1</td>
<td>Type 1 inflammation</td>
</tr>
<tr>
<td>Th17</td>
<td>Type 17 inflammation</td>
</tr>
<tr>
<td>Th2</td>
<td>Type 2 inflammation</td>
</tr>
<tr>
<td>Th22</td>
<td>Type 22 inflammation</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumour Necrosis Factor</td>
</tr>
</tbody>
</table>
Introduction

Atopic dermatitis (AD) arises from a complex interaction between impaired epidermal barrier and environmental exposures to allergens and irritants, resulting in aberrantly activated infiltrating immune cells. Much interest has focused on the immune cells infiltrating AD skin, which mediate the disease. In particular, dense infiltration of activated Th2/Th22 CD4+ T cells has been observed as an early feature of AD exacerbations, especially in acute lesions. These are identified also in non-lesional skin of AD sufferers suggesting a systemic immunodysregulation. This implies that type 2 cytokines play a major role in disease pathogenesis and clinical research has shown the impressive efficacy in AD treatment of a monoclonal antibody therapy targeting IL4Ra which blocks IL-4 and IL13 signalling. However, studies by Gittler et al. first demonstrated that the Th1/Th17 axis is also prominent in chronic AD lesions and correlates with the magnitude of the Th2 signals. Whilst various different T cell pathways have been targeted in clinical trials of AD, the functional effects of the inflammatory pathways on skin keratinocytes have largely been ignored.

Alongside the immune skin infiltrate, spongiosis and keratinocyte hyperplasia are the cardinal features of epidermal changes in AD. In addition to the gene mutation mediated reduction of filaggrin expression, type 2 inflammation also reduces keratinocyte filaggrin expression thereby further damaging the skin barrier. Importantly, beyond their role in maintaining the physical barrier of the skin, keratinocytes also act as innate immune sentinels, and express pattern recognition receptors, ligation of which regulates keratinocyte synthesis of cytokines, antimicrobial peptides (AMPs) and antigen presentation to immune cells.
The role of IL-17 and IL22 cytokines in regulating antimicrobial peptides such as s100 proteins and β-defensins is well established\textsuperscript{17–22} and the importance of these pathways in psoriasis has been validated by clinical demonstration of effectiveness of inhibitory monoclonal antibody therapy\textsuperscript{23,24}; their precise function in AD is less clear. To study the key pathways driving AD, where targeted intervention may prove most fruitful, direct quantitation of the immune signals (e.g. cytokines) can be undertaken. However, as molecular quantitation of the cytokine transcripts does not predict the importance of a specific pathway in AD, it is necessary to study the outcome of the epidermal responses to different inflammatory stimuli to properly define their role. Thus, to characterise the keratinocyte immunophenotype it is necessary to be able to investigate skin transcriptome to a cellular resolution. Single-cell analysis can offer an approach to this question but is limited by the technical challenge of achieving adequate encapsulation of enough cells of interest with minimal transcriptomic disturbance. Here we show that it is possible to employ machine learning to resolve the keratinocyte transcriptomic signal from the non-keratinocyte skin transcriptome, revealing important insights to the pathogenesis of AD.
Methods

Microarray data analysis

Microarray datasets were obtained from the Gene Expression Omnibus (GEO, NCBI) and were analysed from raw data in R and normalised according to platform specifics. For Affymetrix microarray platforms, classic microarray quality control was performed using the Bioconductor arrayQualityMetrics tool and were normalised to obtain expression values by GCRMA methods within the Affymetrix package. Illumina platforms and other technologies were quantile normalised using the lumi or limma Bioconductor packages.

Unsupervised network clustering

The inflammatory skin disease datasets GSE32924, GSE36842 and GSE34248 were processed as above and subsequently merged and batch corrected using the COMBAT tool within the SVA Bioconductor package. Differential gene expression analysis between healthy controls and lesional skin, within disease lesional and non-lesional, and across disease lesional comparisons was conducted using a filtering of Benjamini-Hochberg adjusted p-value >0.05, log(2)-fold difference x1 using the LIMMA Bioconductor package. The expression values of 4620 probset-IDs, corresponding to 3066 unique genes were input into MIRU (now, GraphiaPro) for network analysis. A transcript-to-transcript correlation matrix using Pearson correlation coefficient of $r \geq 0.7$ was created. The resulting network graph was then clustered into groups of genes using the MCL algorithm at an inflation value of 3.1 and minimum cluster size of 10 genes, giving 50 clusters. The gene list for each cluster was interrogated for gene ontology using the web-based analysis tool ToppFun within the ToppGene suite. The REVIGO online tool was used to provide
a single biological process term for each cell-based cluster, selecting that with both the lowest B.H. p-value and a term dispensability of zero.

**Reference cutaneous cell populations datasets**

To curate cell profiles for input as CIBERSORT reference signatures, we collated seven datasets from GEO: GSE36287 (keratinocytes stimulated with IFNα, IFNγ, IL-13, IL-17A, IL-4, TNFα, and unstimulated control), GSE7216 (keratinocytes stimulated with IL-1β, IL-22, IL-26, keratinocyte growth factor (KGF)), GSE34308 (dermal fibroblasts), GSE74158 (skin resident CD4, CD8 and regulatory T cells), GSE4570 (melanocytes), GSE49475 (activated Langerhans’ cells, CD11c+ dermal dendritic cells), and GSE23618 (steady-state Langerhans’ cells). Datasets were normalised separately, and gene expression of sample replicates averaged by mean before combining into a single file for upload to CIBERSORT as a signature genes matrix. Gene replicates are discounted by CIBERSORT in favour of the gene with the highest mean expression across the samples.

**Bulk skin datasets**

Datasets of skin biopsies from inflammatory skin diseases were obtained from GEO. Four AD datasets and two psoriasis datasets were analysed. GSE32924 is a dataset of paired chronic lesional (AD-ChL, n=11), non-lesional (AD-Non, n=11) atopic dermatitis samples, and healthy controls (HH, n=8). GSE36842 is a dataset of acute lesional (AD-AcL, n=7), chronic lesional (AD-ChL, n=7), non-lesional (AD-Non, n=7) skin from atopic dermatitis patients and healthy controls (HH, n=6). GSE58558 is a longitudinal study of ciclosporin treatment of atopic dermatitis patients with biopsies taken at baseline, 2 and 12 weeks of treatment of both chronic lesions (CAL:
baseline, n=16; 2-weeks, n=17; 12-weeks, n=17) and non-lesional skins (ANL: baseline, n=16; 2-weeks, n=16; 12-weeks, n=16). GSE130588 is a longitudinal study of dupilumab or placebo treatment of atopic dermatitis patients with biopsies taken at baseline, 4 and 16 weeks of treatment of both chronic lesions (AD-ChL: baseline-dupilumab, n=26; baseline-placebo, n=25; 4-weeks-dupilumab, n=24; 4-weeks-placebo, n=20; 16-weeks-dupilumab, n=16; 16-weeks-placebo, n=16) and non-lesional skins (AD-Non: baseline-dupilumab, n=22; baseline-placebo, n=20; 16-weeks-dupilumab, n=15; 16-weeks-placebo, n=7) and healthy controls (HH, n=20).

GSE34248 is a dataset of paired lesional (Ps-Les, n=14) and non-lesional (Ps-Non, n=14) psoriatic samples. GSE11903 is a longitudinal study of etanercept treatment of psoriasis patients with biopsies taken at baseline for lesional and non-lesional skin (Ps-Les, n=11; Ps-Non, n=11) and biopsies from lesional skin at 1- (n=11), 2- (n=11), 4- (n=10), and 12 weeks (n=11) of treatment

Running CIBERSORT

Datasets of bulk skin samples were deconvoluted against the reference signature sets using the online version of the CIBERSORT\textsuperscript{28} algorithm. Reference signature files were provided as the signature gene file, while normalised expression data of bulk samples were provided as the mixture file. All run settings were kept at default. Output provided by CIBERSORT was downloaded as a .txt file, where relative abundance of each cellular signature was normalised as a percent of sample.
Results

Gene co-expression analysis of inflammatory skin disease genes reveals immune and keratinocyte involvement.

Firstly, we set out to examine the transcriptomic signals from skin biopsies of lesional, regardless of chronicity, and non-lesional AD and psoriasis (AD-Les, AD-Non, Ps-Les, Ps-Non, respectively). In line with previous reports, unsupervised differential expression analysis identified 4620 genes\(^3,29\). Transcript-to-transcript clustering (GraphiaPro, Pearson r >0.7, MCL=3.1, >10 clustered genes) identified 50 clusters (Figure 1a). Annotation revealed three clusters (11, 14 and 28) encoding immune-related processes such as lymphocyte activation, interferon and cytokine signalling (Figure 1f-h; Supplementary table 1). The relative expression pattern across these clusters showed similar changes in lesions, regardless of disease, and the least expression in healthy skin. Interestingly, AD non-lesional skin showed a prominent defence response (cluster 28), suggesting a subclinical immune alteration in non-lesional AD skin as compared to healthy skin\(^3,30,31\).

Strikingly, clusters 1, 6, 9 and 18 (Figure 1b-e; Supplementary table 1) were enriched in biological processes characteristic for keratinocytes (KCs). Genes in clusters 1 and 18 were most highly expressed in healthy tissue and represented processes of epidermis development and skin development, respectively, suggesting aberrant regulation of these processes in inflammatory skin disease. In contrast, genes in cluster 9 showed strong correlation with expression of immune mediated inflammation genes which suggests that the association may be causal. Cluster-to-cluster gene expression by tissue showed high correlation (Pearson r-square >0.7 (data not shown)). It is notable that genes in cluster 9 include \textit{KRT6A/B, KRT16}, and
the S100 and SERPIN encoding proteins which are known to be involved in epidermal perturbation from inflammation and hyper-proliferative barrier breach\textsuperscript{2,32,33}, indicating that the cross-talk between immune inflammation and keratinocyte function is important for lesion pathogenesis.

**Machine learning resolution of whole skin samples into constituent cellular profiles.**

To resolve from bulk expression data the transcriptomic signatures of keratinocyte responses we utilised a machine learning approach. We trained an algorithm (CIBERSORT\textsuperscript{28,34}) to resolve gene expression profiles of purified cellular populations and tested this on whole tissue transcriptomic data from split skin (epidermis and dermis) to identify the relative proportion of cells. We then utilised laser captured dermal and epidermal regions from both healthy and AD skin\textsuperscript{35} to demonstrate that the algorithm could reliably separate relative proportions of the keratinocyte and fibroblast composition of whole skin in uninflamed and inflamed settings (Figure 2a, b). We expanded this approach by utilising a training set of transcriptomes for melanocytes, resident regulatory, CD4 and CD8 T cells, steady-state and activated Langerhans cells, and CD11c\textsuperscript{+} dermal dendritic cells to increase the resolution of the cellular skin components (Figure 2c).

**In silico sorting reveals prevalence of KC2 and progression of KC17 immunophenotypes during the course of the inflammation in AD.**

To identify sub-populations of keratinocytes showing a molecular response to a specific inflammatory cytokine, we further trained the algorithm to resolve keratinocytes responding to IFN-\(\alpha\) and IFN-\(\gamma\) (KC1), IL-4 and IL-13 (KC2), IL-17A...
(KC17), KGF, IL-26, IL-22, TNF, IL-1β, as well as resting (Figure 3b-e, Supplementary figure S1). These were tracked across 14 healthy (HH), 18 AD non-lesional samples (AD-Non), seven AD acute lesional samples (AD-AcL) and 18 AD chronic lesional samples (AD-ChL) (GSE32924 and GSE36842) and 14 psoriasis non-lesional samples (Ps-Non) and 14 in psoriasis lesional (Ps-Les) to investigate disease-related shift in the transcriptomic programme of these cells (Figure 3b-e). As expected, in AD, a strong KC2 signal was clearly detectable, showing that keratinocytes in chronic lesions were significantly responding to type 2 cytokines in comparison to healthy skin (ANOVA, p<0.05) (Figure 3c). Interestingly, this signal was equally strong in non-lesional skin suggesting that this dysregulation may be systemic in AD.

A keratinocyte interferon programme, KC1, was prominent in lesional and absent from non-lesional psoriatic samples (ANOVA, p<0.0001) (Figure 3b). KC1 was also found in chronic lesions from AD samples, showing a trend of increase from non-lesional and acute lesional stages (Figure 3b), inferring the complex Th2/Th17/Th1 inflammation experienced by the epidermis of chronic AD lesions also alters the transcriptomic programme of keratinocytes. As expected, the KC17 signal was significantly elevated in psoriatic lesions compared to healthy controls and paired non-lesion samples (ANOVA, p<0.0001) (Figure 3d). In contrast to AD, the Th17 signal was absent in both non-lesional psoriatic skin and skin of healthy controls. Compared to healthy controls, AD keratinocytes showed an IL-17 sensing signal in lesional as well as non-lesional skin, but this was only significant for chronic lesions (ANOVA p<0.0001) (Figure 3d). KC17 appeared to dominate in chronic AD lesions, suggesting an evolution in Th17 pathway over time.
**Treatment specific modification of keratinocyte immunophenotype signature.**

Microarray data from psoriasis patients treated with etanercept, showed longitudinal abrogation of KC17 (p<0.0001) underscoring the role of this pathway in psoriasis pathogenesis (Figure 4b). The immunophenotypic signature of response to AD treatment was more nuanced. Despite efficacy and improvement in disease severity scores across the cohort (SCORAD50 improvement), ciclosporin treatment did not reduce the KC2 fraction in either non-lesional or lesional samples (Figure 4c). However, the KC17 fraction in lesional keratinocytes was significantly reduced by ciclosporin (p=0.02) which evolved to show no significant difference from non-lesional skin (p>0.05) (Figure 4d).

To follow the effect of anti-Th2 treatment on keratinocyte transcriptomic responses, we used the GEO microarray dataset GSE130588, where bulk skin microarray data was obtained from AD patients treated with 400mg of dupilumab with biopsies taken at treatment initiation, 4-weeks and 16-weeks on the conclusion of treatment, along with a cohort healthy control. Biopsies were taken from the same lesional site and showed a clinical and transcriptomic improvement in clinical atopy with treatment in established, ongoing lesions. Interestingly, very few of the AD samples, including many of the lesional biopsies, showed a KC17 fraction. Dupilumab treatment demonstrated a dramatic early reduction in the KC2 fraction of lesions at 4 weeks compared to baseline (p=0.0007), which was sustained at 16 weeks (p=0.0002) (Figure 4e). At the start of treatment, AD lesions had an increased KC2 profile compared to healthy controls, which resolved to be comparable by the end of anti-IL-4/IL-13 treatment.
Discussion

Despite our understanding of the role of inflammatory cells in AD, and identification of the different inflammatory signals evident in AD, it is surprising that little attention has been paid to characterising the keratinocyte response in detail. This has perhaps, in part, been due to technical challenges associated with addressing the question. Standard approaches to bioinformatic analysis of transcriptomic studies of AD employ statistical tools to identify differentially expressed genes in lesions versus non-lesions and can utilise gene set enrichment analysis based on the functional annotation of the differentially expressed transcripts to identify cellular processes which are more or less prominent in AD.

However, such a macro view of the AD transcriptome prevents characterisation of individual responses of the various cell-types comprising skin. Widely used approaches to look at individual cell populations include flow cytometry and immunohistochemistry. Despite the routine application of these methodologies, and valuable insights they can provide for a relatively narrow set of markers, they only inform analysis of cell phenotype in a relatively limited way, partly because both of these techniques require a monoclonal antibody label of which a limited panel can be applied to a single sample. Single cell sequencing from whole skin or from flow sorted populations would allow the investigator to undertake detailed characterisation of cell types in AD in a non-hypothesis driven manner. However, as yet, these approaches are limited by cost, and the relatively low number of sequenced cells as a proportion of the whole tissue sample may present a sampling bias. We sought to address cellular analysis within tissue by a different approach. Using machine learning on published datasets, we trained an algorithm to identify
different skin cell types and then cellular responses to different inflammatory responses of interest. We validated this algorithm on transcriptomic studies of microdissected healthy skin and psoriasis which has a known immune pathway dependence, showing that our approach is a powerful method for investigating transcriptomic signatures in skin samples of complex disease.

Applying such optimised machine learning-based analysis to existing datasets of AD disease stages and during treatment has confirmed the constitutive atopic skin phenotype in AD patients. An altered transcriptomic programme in keratinocytes was evident in all samples (lesional and non-lesional) which reflected Th2 sensing by keratinocytes which we termed “KC2”, and similar findings have been reported by others\(^3,5\). Further, we show the immunophenotype shift characteristic of lesion progression modifies keratinocyte profiles to an IL-17 (KC17) and interferon (KC1) sensing phenotype. Thus, we could demonstrate that although acute AD lesions show a strong Th2 signal, and chronic lesions have Th1 and Th17 signals, the Th2-related processes is amplified rather than a switch away from a type 2 cytokine response in chronic lesions.

Effective treatment of psoriasis with ciclosporin showed a reversal of the dominant KC17 profile of lesional skin to that of unaffected skin. However, in AD, remarkably, and despite resolution of skin inflammation as measured by disease severity, ciclosporin did not modify the KC2 profile of lesional skin. Instead, AD disease remission with ciclosporin correlated with loss of the KC17 signal. In contrast to this, with dupilumab, striking loss of the KC2 signal associated with disease remission, whereas the KC17 response was not observed. This underscores the critical role of
the type 2 cytokines in AD and might suggest a strong role for IL-17 pathway in AD pathogenesis; it may also reflect the complex immunophenotype of the disease and potential immune mediator redundancy.

Our analysis to computationally resolve keratinocyte sub-populations by their sensing of immune-related signals does not address AD as a disease driven by epidermal disruption or systemic immune abnormalities. Indeed, we see evidence of some individual variation at a keratinocyte level, particularly of KC2 and KC17 immunophenotypes. We postulate that various environmental factors such as commensal dysbiosis may contribute to individual variation in epidermal sensing by keratinocytes and may regulate the epidermal response, both modifying and modified by immune infiltrate signals. This theory would suggest that perturbation of the immune signal alone may, in some situations, be insufficient to resolve the keratinocyte immunophenotype. Furthermore, such considerations emphasise the importance of characterising the epidermal responses alongside the immune signals in molecular studies of AD.

In summary, in-silico deconvolution of the transcriptional phenotype of AD keratinocytes has revealed two levels of pathology. First, individuals with AD epidermis demonstrate keratinocytes sensing of type 2 cytokines. Secondly, although the IL-4/IL-13 signal becomes enhanced in chronic AD lesions, it appears that induction of an IL-17 response acts as a key switch between acute/chronic AD. This confirms the model of sequential activation of helper T-cell responses across the development and chronicity of cutaneous lesions. Finally, we showed that despite disease resolution with both ciclosporin and dupilumab, ciclosporin treatment
rebalances KC17 subpopulation comparable to normal skin but does not modify type 2 cytokine sensing. Whereas, dupilumab therapy reverses the KC2 dominance in lesions of AD. Taken together, these observations suggest that whilst type 2 cytokines appear to drive the biology of AD, the efficacy of ciclosporin in AD is likely to lie beyond the targeting of T cells with resultant Th17 inhibition and that other pathways modified by this effective therapy should be explored as potential therapeutic targets.
Acknowledgements

We thank our funding bodies the MRC and Wellcome Trust, and the iCASE PhD studentship sponsor, Unilever, for providing the resources to undertake this research. We also thank Rebecca Ginger, formerly of Unilever (Colworth, UK) for previous discussions and contributions.
References


Severity Index. Acta Derm Venereol. 2019;0.


23. Guttman-Yassky E, Lowes MA, Fuentes-Duculan J, Zaba LC, Cardinale I,


Supporting information

Supplementary Data 1.

Co-expression clusters from Figure 1 (50 clusters).

50 gene clusters were identified using co-expression analysis (gene-to-gene expression correlation Pearson $r > 0.7$, and MCL 3.1). The gene lists (first tab of Data E1_co-expression genes and GOBP.xlsx file) was input to ToppGene for gene ontology annotation for biological process meaning (GOBP) (ToppGene results tabs labelled 1-50 in Data E1_co-expression genes and GOBP.xlsx file. See file: Data S1_co-expression genes and GOBP.xlsx

Supplementary Figure S1.

Resolution of whole skin samples into constituent cellular profiles using machine learning.

(a) Deconvolution of keratinocyte signature into profiles representing keratinocyte immunophenotypes across the six skin conditions: healthy (HH), AD non-lesional (AD-Non), AD acute lesional (AD-AcL), AD chronic lesional (AD-ChL), psoriasis non-lesional (Ps-Non), psoriasis lesional (Ps-Les). (Default CIBERSORT settings, training signatures: keratinocytes (unstimulated, stimulated with IFN$\alpha$, IFN$\gamma$, IL-17, IL-1$\beta$, TNF, IL-22, IL26, KGF); fibroblasts; melanocytes; CD4+, CD8+ and regulatory T cells; dermal dendritic cells; and, steady-state and activated Langerhans cells here, supplementary table E4) (b-f) Individual keratinocyte immunophenotypes; (b) unstimulated KC; (c) response to IL-1; (d) response to keratinocyte growth factor (KGF); (e) response to TNF-$\alpha$; (f) response to IL-26 across the five tissue types. Healthy patients (n=14, green), AD non-lesion patients (n=18, red), AD acute lesion patients (n=7, cyan), AD chronic lesion patients (n=18, blue), psoriasis non-lesion
patients (n=14, orange), and psoriasis lesion patients (n=14, magenta). ANOVA Sidak’s multiple test bars below: p>0.05,ns; p<0.05, *; p<0.01, **; p<0.001, ***; p<0.0001, ****. Error bars show mean± SD. (Additional to KC fractions shown in main figure 3b-e).

See file: Supplementary figure S1.pdf

Supplementary Table 1.

Gene ontology annotation of the cell-based co-expression clusters from whole skin microarray data.

Seven cell-based clusters were annotated from gene co-expression analysis (Figure 1A). The clusters were annotated for biological process gene ontology (GOBP) using the ToppFun/ToppGene; the top 200 annotations with FDR Benjamini-Hochberg-corrected p-values <0.05 were selected and collapsed for redundancy using REVIGO. The most significant collapsed annotation from REVIGO was selected as the cluster biological process annotation, with the ToppGene rank shown.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Genes</th>
<th>Biological process</th>
<th>GO term</th>
<th>P-value (BH FDR)</th>
<th>GO term rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>641</td>
<td>Epidermis development</td>
<td>GO:0008544</td>
<td>1.60E-04</td>
<td>1</td>
</tr>
<tr>
<td>Cluster 6</td>
<td>104</td>
<td>Keratinocyte differentiation</td>
<td>GO:0030216</td>
<td>7.14E-05</td>
<td>2</td>
</tr>
<tr>
<td>Cluster 9</td>
<td>68</td>
<td>Skin development</td>
<td>GO:0043588</td>
<td>1.09E-05</td>
<td>2</td>
</tr>
<tr>
<td>Cluster 11</td>
<td>63</td>
<td>Regulation of lymphocyte activation</td>
<td>GO:0051249</td>
<td>1.51E-18</td>
<td>1</td>
</tr>
<tr>
<td>Cluster 14</td>
<td>39</td>
<td>Type I IFN signalling pathway</td>
<td>GO:0060337</td>
<td>8.30E-24</td>
<td>1</td>
</tr>
<tr>
<td>Cluster 18</td>
<td>32</td>
<td>Skin development</td>
<td>GO:0043588</td>
<td>3.02E-07</td>
<td>1</td>
</tr>
<tr>
<td>Cluster 28</td>
<td>18</td>
<td>Defence response</td>
<td>GO:0006952</td>
<td>7.26E-03</td>
<td>1</td>
</tr>
</tbody>
</table>
**Supplementary Table 2.**

Training samples of unstimulated keratinocytes (GSE36287) and skin dermal fibroblasts (GSE34308) with expression reduced to mean expression across replicates.

See file: Table S2.xlsx

**Supplementary Table 3.**

Training samples of unstimulated keratinocytes (GSE36287); skin dermal fibroblasts (GSE34308); CD4+, CD8+ and regulatory T cells (GSE74158), melanocytes (GSE4570), activated Langerhans cells and CD11C+ dermal dendritic cells (GSE49475); and steady state Langerhans cells (GSE23618), with expression reduced to mean expression across replicates.

See file: Table S3.xlsx

**Supplementary Table 4.**

Training samples of keratinocytes (unstimulated, stimulated with IFNα, IFNγ, IL-17, IL-1β, TNF, IL-22, IL26, KGF (GSE36287, GSE7216)); skin dermal fibroblasts (GSE34308); CD4+, CD8+ and regulatory T cells (GSE74158), melanocytes (GSE4570), activated Langerhans cells and CD11C+ dermal dendritic cells (GSE49475); and steady state Langerhans cells (GSE23618), with expression reduced to mean expression across replicates.

See file: Table S4.xlsx
Figure Legends

Figure 1 legend

Unsupervised co-expression analysis of whole skin from healthy, atopic and psoriatic skin. (a) Transcript-to-transcript clustering of 4620 differentially expressed genes of atopic lesional (AD-Les), atopic non-lesional (AN-Non), psoriatic lesional (Ps-Les) and psoriatic non-lesional (Ps-Non) compared to healthy control (HH) (FDR p-value <0.05, log fold-change <-2/>2). Gene to gene co-expression correlation of >0.7 Pearson were retained for Markov clustering using an inflation value of 3.1. (b-h) The top 50 clusters were annotated for biological process, of which seven clusters were identified as cell-based. Average expression of the all the genes in cluster shown per phenotype. ANOVA Sidak’s multiple test bars below: p>0.05, ns; p<0.05, *; p<0.01, **; p<0.001, ***; p<0.0001, ****. Error bars show mean± SD.

Figure 2 legend

Machine learning resolution of whole skin samples into constituent cellular profiles. (a) Identification of key cellular signatures (fibroblasts: dark grey and keratinocytes: light grey) in microdissected samples from dermis and epidermis by machine learning (healthy dermis, n=6; healthy epidermis, n=10). (b) Inflammation/disease status (HH, healthy; AD-Non, AD non-lesional sample; AD-ChL, AD chronic lesion sample) of dermis or epidermis does not affect the correct deconvolution of cellular components within the bulk disease tissue (dermis: HH, n=6; AD-Non, n=5; AD-ChL, n=5. Epidermis: HH, n=10; AD-Non, n=5; AD-ChL, n=5). (a,b) (default CIBERSORT settings: 100 permutations, kappa = 999, q-value = 0.3, number of barcode genes 50-150. Training signatures: GSE36287, unstimulated keratinocytes; GSE34308, skin fibroblasts, Supplementary table E2) (c) Whole skin
from healthy controls and AD patients (non-lesional, AD-Non; acute and chronic lesional, AD-AcL and AD-ChL, respectively) resolved into relative fractions of cutaneous cell populations of 9 transcriptomic signatures; keratinocytes; fibroblasts; melanocytes; CD4+, CD8+ and regulatory T cells; dermal dendritic cells; and, steady-state and activated Langerhans cells). The mean percentage of each of the signatures is shown relative to the remaining signatures. Healthy controls (n=14), and AD patients (non-lesional, n=18; acute lesional, n=7; chronic lesional, n=18). (default CIBERSORT settings. Training signatures: Supplementary table E3).

**Figure 3 legend**

**Resolution of whole skin samples into constituent cellular profiles using machine learning.** (a) Deconvolution of keratinocyte signature into profiles representing keratinocyte immunophenotypes across the six skin conditions: healthy (HH), AD non-lesional (AD-Non), AD acute lesional (AD-AcL), AD chronic lesional (AD-ChL), psoriasis non-lesional (Ps-Non), psoriasis lesional (Ps-Les). (Default CIBERSORT settings, training signatures: keratinocytes (unstimulated, stimulated with IFNα, IFNγ, IL-17, IL-1β, TNF, IL-22, IL26, KGF); fibroblasts; melanocytes; CD4+, CD8+ and regulatory T cells; dermal dendritic cells; and, steady-state and activated Langerhans cells here, Supplementary table E4) (b-e) Individual keratinocyte immunophenotypes; (b) KC1 (response to IFNα, IFNγ); (c) KC2 (response to IL-4, IL-13); (d) KC17 (response to IL-17), (e) KC22 (response to IL-22); remaining KC fractions shown in Supplementary figure E1. Healthy patients (n=14, green), AD non-lesion patients (n=18, red), AD acute lesion patients (n=7, cyan), AD chronic lesion patients (n=18, blue), psoriasis non-lesion patients (n=14, orange), and psoriasis
lesion patients (n=14, magenta). ANOVA Sidak’s multiple test bars below: p>0.05, ns; p<0.05, *; p<0.001, **; p<0.0001, ****. Error bars show mean± SD.

Figure 4 legend
Response of disease-related keratinocyte fractions, KC2 and KC17, to treatment of psoriasis and AD. (a, b) KC2 (a) and KC17 (b) in whole skin from psoriasis patients undergoing etanercept treatment (non-lesional (week 0 only, n=11, orange), and lesional (weeks 0, 1, 2, and 12, n=11; week 4, n=10, magenta) [GSE11903 (Default CIBERSORT settings, training signatures: as Figure 3, Supplementary table E4)]. (c, d) KC2 (c) and KC17 (d) during a time course of ciclosporin treatment at baseline (AD-Non (red), n=16; AD-ChL (blue), n=16), mid-treatment (2 weeks) (ANL, n=17; CAL, n=16), and end of treatment at 12 weeks (ANL, n=17; CAL, n=16). [GSE58558 (Default CIBERSORT settings, training signatures: as Figure 3, Supplementary table E4)] (e, f) KC2 (e) and KC17 (f) fractions in healthy (green), AD non-lesional (red) and AD lesional (blue) AD samples during a time course of dupilumab treatment at baseline (HH, n=20; AD-Non, n=42; AD-ChL, n=51), mid-treatment (4 weeks) (AD-ChL, dupilumab-treated n=24), and end of treatment (16 weeks) (AD-ChL, dupilumab-treated n=18). [GSE130588 (Default CIBERSORT settings, training signatures: as Figure 3, Supplementary table E4)].
Figure 1

Cutaneous cell-type clusters (b-e)

Immune cell-type clusters (f-h)
Figure 3

(a) Keratinocyte response, KC1 (IFNα, IFNγ)

(b) Keratinocyte response, KC2 (IL-4, IL-13)

(c) Keratinocyte response, KC17 (IL-17)

(d) Keratinocyte response, KC22 (IL-22)

(e) Keratinocyte response, KC1 (IFNα, IFNγ)

Legend:
- Healthy
- AD, Non-lesion
- AD, Acute Lesion
- AD, Chronic lesion
- Psoriasis, Non-lesion
- Psoriasis, Lesion

Statistical tests:
- ANOVA
- ns
- *p < 0.05
- **p < 0.01
- ***p < 0.001
- ****p < 0.0001

Cellular fraction, %

Keratinocytes, unstimulated
Keratinocytes, IFNα & IFNγ (KC1)
Keratinocytes, IL4 & IL13 (KC2)
Keratinocytes, IL17 (KC17)
Keratinocytes, IL22 (KC22)
Keratinocytes, IL26
Keratinocytes, KGF
Fibroblasts
Melanocytes
CD4 T cells
CD8 T cells
Treg cells
LC, steady state
LC, activated
DDC
Treg cells
CD8 T cells
CD4 T cells
Melanocytes
Keratinocytes, KGF
Keratinocytes, IL26
Keratinocytes, IL22 (KC22)
Keratinocytes, TNF
Keratinocytes, IL1
Keratinocytes, IL17 (KC17)
Keratinocytes, IL4 & IL13 (KC2)
Keratinocytes, IFNγ & IFNα (KC1)
Keratinocytes, unstimulated
Figure 4

Keratinocyte response, KC2 (IL4, IL-13)

(a) Ps-Non

Ps-Les

(b) Keratinocyte response, KC17 (IL-17)

Ps-Non

Ps-Les

(c) AD-Non

AD-ChL

(d) Time course

Week 0, week 2, week 12

Week 0, week 2, week 12

(e) AD-Non

AD-ChL

(f) Time course

Week 0, week 2, week 12

Week 0, week 2, week 12

(g) Healthy

AD-Non

AD-ChL

(h) Time course

Week 0, week 4, week 16

Week 0, week 4, week 16

Note: All rights reserved. No reuse allowed without permission.